JIŘÍ JANATA

Principles of
Chemical
Sensors

Second Edition

Springer

Principles of Chemical Sensors

Jiří Janata

Principles of Chemical Sensors

Second Edition

Springer

Dr. Jiří Janata
Georgia Institute of Technology
School of Chemistry & Biochemistry
Atlanta GA 30332-0400
USA
jiri.janata@chemistry.gatech.edu

Illustrations by Courtney N. Phillips

Printed on acid-free paper

To the memory of my teachers

Preface

Do not learn the tricks of the trade, learn the trade

I started teaching graduate courses in chemical sensors in early 1980s, first as a one-quarter (30 h) class then as a semester course and also as several intensive, 4–5-day courses. Later I organized my lecture notes into the first edition of this book, which was published by Plenum in 1989 under the title *Principles of Chemical Sensors*. I started working on the second edition in 2006. The new edition of *Principles of Chemical Sensors* is a teaching book, not a textbook. Let me explain the difference.

Textbooks usually cover some more or less narrow subject in maximum depth. Such an approach is not possible here. The subject of chemical sensors is much too broad, spanning many aspects of physical and analytical chemistry, biochemistry, materials science, solid-state physics, optics, device fabrication, electrical engineering, statistical analysis, and so on. The challenge for me has been to present uniform logical coverage of such a large area. In spite of its relatively shallow depth, it is intended as a graduate course. At its present state the amount of material is more than can be covered in a one-semester course (45 h). Two one-quarter courses would be more appropriate. Because of the breadth of the material, the sensor course has a somewhat unexpected but, it is hoped, beneficial effect. It cannot pass as a textbook in any single area, yet it links different subjects that are not covered by the specialized textbooks. In that respect it serves as a review of many areas of chemistry, engineering, and physics.

In my experience, even the first-year graduate students taking this course do not have sufficient knowledge of diverse areas. For that reason, appendices have been included as a quick reference. It is strongly recommended that the non-chemistry students review Appendices A and B, before starting on Chapters 1–3, and the non-physically oriented students review Appendices C and D before Chapter 4 and further. The reader will find discussion on why some sensors work or do not, or even cannot, work. It is not a book that will direct the student to the specific type of sensor for a specific compound. There are plenty of modern search engines that allow retrieval of such information. This book is about principles.

Throughout the book "Thought Experiments" and analogies, such as equivalent electrical circuits, have been used to explain certain working principles. A logical flow in derivations of equations and/or explanation of concepts has been preserved. It is recommended that the instructor prepares his or her own lecture notes and use the book mainly as supporting lecture material. Testing of students' knowledge, otherwise known as exams, is an integral part of any course. I tried to avoid using conventional numerical problem sets because they tend to promote mechanical use of concepts, rather than to promote their understanding. There are plenty of problem sets that can be found in specialized textbooks. The instructor can find and "borrow" them, if necessary.

Quoting the great medieval Bohemian educational reformer John Amos Comenius (1592–1670) that "the best way to learn is from mistakes," I followed his advice in my book. There are many examples in the chemical sensor literature that are not correct, or that cannot even work, for principal reasons. Arguing with them one-by-one inevitably leads to endless "battles with windmills," with referees, authors, and editors. As a sole author of this book I am taking advantage of being able to express my own unchallenged opinions on some such controversial concepts and ideas. It gives me an opportunity to explain why I think that something cannot work rather than just declare that "I am right and the author is wrong." In order to take full advantage of that position I have added to this edition a section called "Food For Thought" at the end of each chapter. Examples of somewhat "provocative" questions, which may not have an obvious answer, but which require thought, can be found in that section. Most have been inspired by published articles and are quoted without attribution. Because publication of such impossible schemes is a never-ending process, a "Food-For-Thought" web site has been established as a "living sensor blog." New FFTs will be added to it periodically. For the time being I will be the custodian, but the readers are invited to submit their entries to the FFT web site. The link can be found on my personal web site, which is at present: www.chemistry.gatech.edu/faculty/Janata/.

The bulk of this book was written during my sabbatical leave at the ETH Zürich, Switzerland and at the Weizmann Institute of Science in Rehovot, Israel. I am eternally grateful to my hosts Professor Andreas Hierlemann (ETHZ) and Professor Israel Rubinstein (WIS), respectively, for providing advice, encouragement, and a stimulating environment in which I could live like a monk and just write and write. That has been the relatively easy part. Multiple level editing, organization, and management of the book proved to be the much harder part of this project. Fortunately, my former student, Liz Thompson, took over this responsibility and lent her energy and considerable expertise to see this project go to completion. There are no words that could adequately express my gratitude for what she has done. Finally, the students in the Spring 2008 Chemical Sensor class proofread and polished the entire manuscript. One of these students, Courtney Phillips, found my original illustrations to be "less than adequate" and offered to add a more artistic flair to them. I acknowledge and thank all of them also for their help.

Atlanta,   *Jiří Janata*
October 2008

Contents

Chapter 1
Introduction to Sensors

1.1 Introduction

Chemical sensing is part of an information-acquisition process in which an insight is obtained about the chemical composition of the system in real-time. In this process, an amplified electrical signal results from the interaction between some chemical species and the sensor. Generally, the interaction consists of two steps: *recognition* and *amplification*. One common example is the measurement of pH with a glass electrode (Fig. 1.1).

The interaction of the hydronium ion with the electrode is highly specific, but the power density in the primary interaction is very low, on the order of pW cm^{-2}. Any attempt to draw current (i.e., obtain a signal) from such an electrode would result in the distortion of the signal. In other words, the source of the signal (electrode) requires a high-input impedance amplifier (pH meter) in order to obtain the information in useful undistorted form. Thus, the recognition (selectivity) is provided by some chemical interaction, whereas the amplification must be provided by some physical transducer. There are some exceptions, for example, enzymatic reactions. Here, the high selectivity of the enzyme binding for a given substrate, combined with the catalytic properties of the enzyme, represents an amplification step in itself.

The coupling of the chemically selective layer to the physical part of the sensor is very important. As we show later, it can have a profound effect on the overall performance of the sensor. In some cases, the highly selective primary interaction can be destroyed by the use of an improper transduction mechanism. The subsequent manipulation of the signal can be done in many different ways and with varying degrees of sophistication. It can be displayed in analog form, subtracted from the reference signal and displayed as a difference, or it can be digitized and processed statistically. Such processing can be done within the physical boundary of the sensor itself ("integrated" or "smart sensors") or it can be carried out in a separate processor.

With so many different sensors described in the literature one may ask: which one is the best? Of course, there is no simple answer, because there are so many
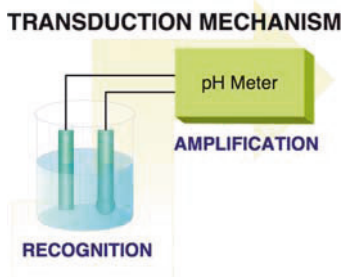
Fig. 1.1 Transduction in chemical sensing consisting of two basic steps: recognition and amplification

different sensing situations and criteria that must be considered. Instead, a general precept must serve: the best sensor is the one that will do the job, at a cost which justifies its use. The cost must be viewed in terms of money, time, and ease of use. These parameters are determined by the characteristics of the sensor.

So, let us first review common sensor terminology. There is one characteristic which stands apart from the others: *ruggedness*. It can be described as the ability of the device to maintain its performance specifications even under adverse operating conditions. In a pragmatic sense, it translates to "reliability." Ruggedness may be assessed in either physical or chemical terms. A physically rugged device is able to withstand mechanical shock, vibrations, and other mechanical stresses. Chemical ruggedness has a slightly more subtle meaning. It is related to selectivity, and it means that the output of the sensor is unaffected by the unforeseen chemical changes in the operating environment.

Ruggedness is usually the most critical criterion in the selection of a sensor for a given application. Rugged sensors are usually the most commercially successful, because they are reliable or because they are the only ones that can provide information under the conditions in which other sensors would not be able to operate.

Generally speaking, we can distinguish two types of interactions between the chemical species and the sensor: a *surface interaction* in which the species of interest is adsorbed at the surface, and a *bulk interaction* in which the species of interest partitions between the sample and the sensor and is absorbed. The classification of the interaction as either surface or bulk is relative with respect to the size of the species. It is the case of "chicken and chicken wire." Obviously, a chicken wire fence is impervious to chickens, but presents no barrier whatsoever to mosquitoes. Similarly, large molecules, such as proteins, may adsorb at the surface of the sensor layer, whereas smaller ions can penetrate and absorb in the bulk.

Reversibility in the context of chemical sensing means that the response follows concentration changes, both up and down. It does not have the usual thermodynamic meaning, despite the fact that a decrease of free energy is always the driving force in all interactions. Thus, sensors can be either thermodynamically reversible, as with ion-selective electrodes, or thermodynamically irreversible, as with enzyme electrodes; if, however, they respond to a step up or a step down in the concentration

of the substrate, they are said to be "reversible" in the context of sensing. Like-wise, detectors can be reversible or irreversible, but dosimeters are nearly always irreversible.

Interaction of a chemical species (X) with sensor (S) can be described by the equilibrium:

$$X + S \underset{k_r}{\overset{k_f}{\rightleftharpoons}} SX \tag{1.1}$$

The rates of the forward (k_f) and reverse (k_r) reactions together with the mass trans-port parameters of the species involved in the transduction mechanism are important for the response of the sensor. Introducing reaction rates into the definition of the equilibrium constant introduces the notion of time. Thus, for the same value of K we can have fast and slow, forward and reverse reactions, and therefore fast or slow equilibrium. The equilibrium constant (K) is expressed in terms of activities.

$$a = fC \tag{1.2}$$

where f is the activity coefficient and C is concentration. The equilibrium constant is then

$$K_X = \frac{a_{SX}}{a_S a_X} = \frac{k_f}{k_r} \tag{1.3}$$

Here a_{SX} is the activity of the bound species and a_X and a_S are the activities of the species in the sample and of the binding site in the sensor, respectively. For the purpose of this discussion, the binding site can be thought of as a defined but separate component of the selective layer, such as in heterogeneous selective layers, or it may be a specific part of the uniform matrix, as in homogeneous selective layers. (More on the origins of selectivity are discussed later.) The free energy of interaction for reaction depicted in (1.1) is

$$0 = \Delta G = \Delta G^0 + RT \ln \frac{a_{SX}}{a_S a_X} \tag{1.4}$$

It follows from (1.4) that upon the change of sample activity, the interaction of the species with the sensor will take place if the change of the standard free energy ΔG^0 is negative. If the binding equilibrium constant is too high ($K > 10^4$; $\Delta G^0 < -23\,\mathrm{kJ\ mole^{-1}}$), the reaction will be, from a practical point of view, nearly irreversible, and the device will respond in a nonequilibrium manner, as a dosimeter.

1.2 General Response Curve

The measurement or "reading" of a sensor is generated by the change in some phys-ical parameter, as a result of some chemical stimulation. This is called the sensor's *response*. Of course, the exact type of physical change depends on both the sen-sor and the sensing environment. For the purposes of the following discussion, we describe response as a general phenomenon.
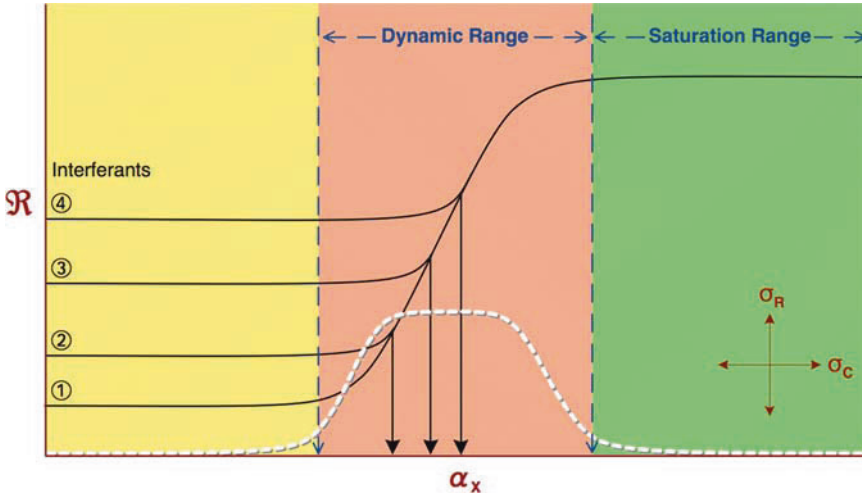
Fig. 1.2 Defined regions of the general response curve

The response of a sensor to the primary species (X) is described by the *response curve* (Fig. 1.2). Besides the analyte (X), many other interfering species (i) also interact with, and competitively bind to, the binding sites in the selective layer. Because such interactions are non-specific, the sites occupied by the interferants are expressed as the sum (Σa_{Si}). Let us assume that only the occupied sites (a_0) in or at the selective layer result in the output signal from the sensor. The total available activity of binding sites (a_{ST}) in the layer is

$$a_{ST} = a_{SX} + a_{si} + a_s \tag{1.5}$$

The activity of unoccupied sites is a_S and activity of the occupied sites giving rise to the signal is

$$a_0 = a_{SX} + \sum_i a_{Si} \tag{1.6}$$

The binding equilibrium for a single interferant (i) is analogous to that of the analyte (see (1.2))

$$K_i = \frac{a_{Si}}{a_S a_i} \tag{1.7}$$

From (1.2), (1.5), (1.6), and (1.7) we can derive the activity of occupied sites:

$$a_0 = a_{ST} \frac{K_X a_X + \sum_i K_i a_i}{K_X a_X + \sum_i K_i a_i + 1} \tag{1.8}$$

Dividing the right-hand side of (1.8) with the value of the binding constant for the analyte (K_X) gives

$$a_0 = a_{ST} \frac{a_X + \sum_i K'_i a_i}{a_X + \sum_i K'_i a_i + 1/K_X} \tag{1.9}$$

In this equation, $K'_i = K_i/K_X$ is the selectivity coefficient. It applies to each interferant individually. The smaller the value of K'_i, the more selective the layer is to the analyte (X) than to the interferant species (i). Equation (1.9) defines three operating regions on the general response curve.

1. At a high activity of the analyte or a high level of the interferants, that is, at $a_X \gg (\sum_i K'_i a_i + 1/K_X)$ and/or $(\sum_i K'_i a_i \gg a_X + 1/K_X)$, all of the available binding sites are occupied by one or the other species. This is the saturation limit of the sensor, for which

$$a_0 = a_{ST} \tag{1.10}$$

   Under these conditions, the sensor does not respond to the analyte X.
2. Below the saturation region lies the dynamic range, for which $\sum_i K'_i a_i \ll a_X \ll 1/K_X$. Here, a fraction of the sites is occupied by X and their activity is linearly dependent on the activity of X in the sample.

$$a_0 = a_{ST} K_X a_X \tag{1.11}$$

3. In the interference region, the activity of the analyte and/or its binding constant are low, that is, $a_X \ll (\sum_i K'_i a_i + 1/K_X)$, and the binding sites are occupied mostly by the interferant. In this region, the sensor again does not respond to the analyte. From (1.9), we have

$$a_0 = a_{ST} \frac{\sum_i K'_i a_i}{\sum_i K'_i a_i + 1/K_X} \tag{1.12}$$

The intercept of the dynamic range (1.11) and of the interference region (1.12) defines the detection limit. This can be expressed for multiple interferants as

$$K_X(a_X)_{d.l.} = \frac{\sum_i K'_i a_i}{\sum_i K'_i a_i + 1/K_X} \tag{1.13}$$

or

$$(a_X)_{d.l.} = \frac{\sum_i K'_i a_i}{\sum_i K_i a_i + 1} \tag{1.14}$$

For a low concentration of a weakly bound single interferant $(K_i a_i \ll 1)$, the selectivity coefficient is given by

$$K'_i = K_i/K_X = \frac{(a_X)_{d.l.}}{a_i} \tag{1.15}$$

(The subscript "*d.l.*" indicates that we are talking about the activity of the analyte at the detection limit.) Equation (1.15) may be used to evaluate the selectivity coefficient for that interferant.

The dynamic range (*D.R.*) is defined as the difference of analyte activities at the saturation limit and at the detection limit. From (1.10) and (1.11), we obtain the activity of analyte X at the saturation limit $[(a_X)_{sl}]$.

$$(a_X)_{s.l.} = \frac{1}{K_X} \tag{1.16}$$

The dynamic range is then derived from (1.15) and (1.16) as

$$D.R. = \frac{1}{K_X} - K_i' a_i \tag{1.17}$$

As the value of term $K_i' a_i$ increases, the dynamic range becomes narrower and narrower until – for a high activity of interferants – it vanishes completely, and the sensor does not respond to the analyte anymore (Fig. 1.2).

At $K_i' a_i \gg 1/K_X$, the sensor again reaches the saturation limit ($a_0 = a_{ST}$), but now most of the sites are occupied by the interferant. The general strategy for achieving optimum sensor design follows from (1.15) and (1.17). We want a low value of the selectivity coefficient and low activity of the interferant. The selectivity coefficient is the ratio of the binding constants of the interferant and of the analyte. Therefore, we want to achieve a low value by decreasing the binding constant of the interferant, not by increasing the value of the binding constant of the analyte, because the first term in (1.17) would decrease the dynamic range.

The output signal from a sensor depends on the transduction function ($\Re_f$). It is related to the occupancy of the binding sites by

$$E_{out} = \Re_f(a_0) \tag{1.18}$$

Although the detection limit depends on the activity of interferants in the sample (i.e., on the *application*), the saturation limit is a property of the sensor itself. This means that the dynamic range is affected primarily by the saturation limit, and depends on the total number of available binding sites. It can be said that the interferants and the selectivity coefficient limit the dynamic range "from the bottom up," whereas the concentration of binding sites limits it "from the top down." For example, a sensor with a detectable minimum density of 10^9 molecules cm^{-2}, and maximum density of 10^{14} sites cm^{-2} (corresponding to $(1\,nm)^2$ molecule^{-1}), has a theoretical dynamic range of 7 decades. How is it then possible that some sensors, for example, the glass electrode, have a dynamic range that extends over 30 decades? There are several reasons.

First, the number of binding sites is much larger for absorption than for adsorption where the maximum surface density is typically 10^{14} sites cm^{-2}. If the effective thickness of the selective layer is $1\,\mu m$ and a bound molecule occupies a cube $(1\,nm^3)$, the maximum number of binding sites increases from 10^{-10} mol cm^{-2} for adsorption up to 10^{-7} mol cm^3 for absorption.

Second, and more important for extending the dynamic range, is the availability of different binding sites with multiple binding constants $(K_1, K_2 \ldots K_n)$, for the species X. In that case, the response will include a contribution from all of the binding events. Ignoring for the moment the interferants, according to (1.11) the activity of the bound species is

$$a_{SX} = a_X \left(K_1 a_{T,1} + K_2 a_{T,2} + \ldots + K_n a_{T,n} \right) \qquad (1.19)$$

where $(a_{T,1}, a_{T,2}, \ldots)$ are the total activities of the individual binding sites. If the partial binding processes do not affect each other, the overall response has the functional relationship $\Re_f$ which applies to all. The output signal is then

$$E_{\text{out}} = \Re_f a_X \left(K_1 a_{T,1} + K_2 a_{T,2} + \ldots + K_n a_{T,n} \right) \qquad (1.20)$$

Such a situation exists within the hydrated layer of a glass electrode, where there exists a multiplicity of binding sites for the hydrogen ion, provided by the statistically large number of lattice defects and the high level of doping impurities. Each site yields an individual dynamic range which then overlaps with the next one, yielding an overall span much greater than would correspond to any single type of binding site with only one equilibrium constant. In the glass electrode, the overall response function is

$$\Re_f = \frac{RT}{zF} \ln a_{SX} \qquad (1.21)$$

In general, a sensing surface has only one type of binding constant, and consequently has only a limited theoretical dynamic range.

The third situation would lead to an extended dynamic range: the change of the sorption mechanism. Appendix A gives several examples of adsorption isotherms. Which isotherm or isotherms apply in any given situation depends on the composition of the selective layer and on the concentration range. Thus, for example, at a low concentration of analyte, the Langmuir isotherm governs the interaction; at a higher concentration, when all available binding sites will have been occupied, a BET isotherm may apply. Also, interaction with specific binding sites may be replaced by interaction with the matrix, and the sorption then follows the simple Henry's law.

The location of the sites is also important from the point of view of the operational characteristics. Sensors whose response depends on bulk interactions are based on the partitioning process, which is governed by the Gibbs equation (A.20). For a two-phase equilibrium this is

$$0 = \Delta G = (\mu_x dn_x)_{\text{sensor}} + (\mu_x dn_x)_{\text{sample}} \qquad (1.22)$$

where X is the chemical species that partitions between the sample and sensor phase and μ_x is the chemical potential of species X. At equilibrium, the number of moles (dn_X) crossing the interface in each direction must be the same (but of the opposite sign):

$$(\mu_x dn_x)_{\text{sensor}} = -(\mu_x dn_x)_{\text{sample}} \qquad (1.23)$$

Thus, the chemical potentials of species X in the two phases must also be equal.

$$(\mu_x)_{\text{sensor}} = (\mu_x)_{\text{sample}} \tag{1.24}$$

This is the general condition of the equilibrium partitioning process. As we show later, it applies to both electrically neutral and electrically charged species. The chemical potential of species X in a phase (gas, solid, or solution) is (c.f. (A.16), (A.19), and (A.25))

$$\mu_X = \mu_X^0 + RT \ln a_X \tag{1.25}$$

Thus, sensors based on absorption (phase equilibrium) measure activity and, if only one type of sorption mechanism is involved, their response is logarithmic.

The condition of the general adsorption equilibrium, which does not assume the existence of high affinity binding sites, is again described by the equality of the chemical potential of the species in the sample phase μ_X and at the surface μ_{SX}. The general adsorption equilibrium has the form of (1.1) and the equilibrium constant (K) can be expressed in a similar manner as above:

$$X + S_{\text{ads}} \leftrightarrow SX_{\text{ads}} \tag{1.26}$$

$$(a_X K_X)_{\text{ads}} = (a_{SX})_{\text{ads}} \tag{1.27}$$

Here SX_{ads} designates the surface adsorbed species. The surface activity is related to the bulk activity through some form of adsorption isotherm.

Throughout this section, we have been using activity in order to emphasize the thermodynamics of the interaction between the analyte and the selective layer. In reality, only a few sensors, notably ion selective electrodes, measure activity. What, then, is actually measured in reality: activity or the concentration of the species? We can take an uncompromising view on this question, and state that all chemical interactions are governed by the activities, which can be equated with concentrations only under special circumstances. In gases, where the interactions between gas molecules are relatively weak, the activities (more appropriately, fugacities) can be equated with concentrations (partial pressures). Another "safe" approximation of activity to concentration is commonly made in dilute aqueous solutions in which the ionic strength does not exceed approximately 1 mM. For electrically neutral species, the concentration of other solutes can be even higher. The immediate consequences of making this assumption without proper justification are the nonspecific interferences due to the variation of the activity coefficient, which translates to a loss of the information content in the signal.

There are, however, sensors in which the output can be justifiably related to concentration, even under conditions where activity and concentration are substantially different. These are the sensors whose output is governed by the spatial gradient or the time-change of concentration, that is, in steady-state sensors for which the output signal E_{out} is

$$E_{\text{out}} = \Re \left(\sum_i K_i \frac{dC_i}{dx} \right) \tag{1.28a}$$

or

$$E_{\text{out}} = \Re\left(\sum_i K_i \frac{dC_i}{dt}\right) \tag{1.28b}$$

If we differentiate the activity (1.2), in distance or in time, we obtain:

$$\frac{da}{dx} = \frac{df}{dx}C + f\frac{dC}{dx} \approx f\frac{dC}{dx} \tag{1.29}$$

or

$$\frac{da}{dt} = \frac{df}{dt}C + f\frac{dC}{dt} \approx f\frac{dC}{dt} \tag{1.30}$$

The changes of activity coefficients over short distances and/or short time periods are almost always close to zero and can be neglected. The changes of activity coefficients are also negligible in the mass transport sensors (e.g., amperometric); there, too, concentration may be used in place of activity. Generally speaking, only the equilibrium sensors with partitioning or adsorption mechanisms depend on activities rather than on concentrations.

A major confusion in the analytical literature, particularly in the sensing literature, is the improper use of "sensitivity" and "detection limit" as interchangeable terms. The *sensitivity* of the sensor is the slope of its response in the dynamic range whereas the *detection limit* is the minimum detectable response, usually above three times the standard deviation (i.e., $>3\sigma$). Because it is evaluated experimentally, the sensitivity of the sensor has two sources of error: the standard deviation of the response (σ_R) and the standard deviation of the concentration (σ_C). The sensitivity (m) of a sensor is the first derivative of its response, with respect to the concentration of the analyte, at any concentration ($C_{X,i}$) on the response curve. Therefore, the sensitivity goes from zero in the interference region, through a maximum in the dynamic range, to zero in the saturation region (Fig. 1.2).

$$m = \left(\frac{dE_{\text{out}}}{dC_X}\right)_{C_{X,j}} \tag{1.31}$$

It has been proposed (Barsan et al., 1999) that analytical sensitivity (γ) be defined as

$$\gamma = \frac{m}{\sigma_R} \tag{1.32}$$

Furthermore, because

$$\sigma_R = m\sigma_s \tag{1.33}$$

it is possible to compare the sensitivity of different sensors using the relationships from (1.32) and (1.33) (Fig. 1.2).

$$\gamma = \frac{1}{\sigma_C} \tag{1.34}$$

Ideally, sensors should behave reversibly in response to the changes of the analyte. If they do not, we talk about sensor hysteresis. This phenomenon can be overlooked

if the response of the sensor is experimentally evaluated only in the direction of increasing concentrations. This can be a problem for the testing of sensors in liquid media. Thus, it is important to evaluate it also for dilutions. A convenient way to do this is to use flow injection analysis as the means of preparing progressively dilute samples.

1.3 Origins of Selectivity

Selectivity is the single most important general issue in chemical sensing. It determines the usefulness of the sensor for any given application. It can be defined as the ability of a sensor to respond primarily to only one species (analyte) in the presence of other species. Its role can be best explained by examining the response curve (Fig. 1.2). In this graph, several responses of the sensor to increasing concentration of the analyte are plotted for increasing concentration of interfering species. As their concentration increases, the detection limit shifts to the higher concentration of the primary species, and the dynamic range becomes narrower and narrower. In the most extreme case it vanishes, and the sensor becomes entirely nonresponding.

Obviously, one of the most important objectives is to design the selective layer in such a way that it has a wide dynamic range and as low a detection limit as possible. It means that, in the ideal case, the layer should completely reject any interfering species, and respond exclusively to the analyte. In a real analytical situation, the nature of interferences is rarely known a priori. For practical reasons, it is desirable to design the selective layer to be as specific as possible. Various strategies for doing that are outlined in the following sections. It is also important to mention here that the selectivity of the sensor response is affected by the correct combination of the recognition mechanism (provided by the selective layer) with the amplification, transduction principle. That aspect is also discussed in individual types of sensors.

There are two general classes of selectivity: equilibrium-based and kinetically based. Both types depend on specific interactions of the analyte molecule with the selective layer. In addition, there is also physical selectivity based on the highly specific interaction of the molecule with the electrostatic and electromagnetic fields. The topic of selectivity is broad enough to deserve its own chapter. We examine issues pertaining to selectivity in Chapter 2.

Food for Thought #1

Irreversible Binding

Given the very high binding constant of some natural binding sites (e.g., DNA, IgG) propose a strategy for optimized utilization of such receptors in direct biosensors.

1. How would you achieve reversibility and at the same time extend the dynamic range of such "designer" biosensors?

Multiple Binding Sites

Both the imine and the secondary amine nitrogens in polyaniline can be protonated by weak carboxylic acids of different pK values. Solid layers prepared from such doped polyanilines have different selectivity for interaction with gases. It is partially derived from the interaction of the analyte with the doping anion.

2. Would you improve the dynamic range if you doped a PANI layer with a mixture of those acids? If not, what would be a better approach?
3. What would happen if there were a "gap" in the range of binding constants, for example, if binding constants between 10^{-6} and 10^{-9} were missing?

Symbols

a	Activity
C	Concentration
E_{out}	Output signal (voltage)
F	Faraday's constant $(96{,}485\,C\,mol^{-1})$
f	Activity coefficient
ΔG^0	Standard free energy
K	Equilibrium constant
K_i	Binding constant for interferant i
K'_i	Selectivity coefficient for interferent i
k	Reaction rate
m	Sensitivity
n	Number of moles
R	Gas constant $(8.314\,J\,K^{-1}\,mol^{-1})$
$\mathfrak{R}_f$	Transduction function
T	Temperature
z	The charge of a species
μ	Chemical potential
σ	Standard deviation of the response
γ	Analytical sensitivity

Reference

Barsan, N., Stetter, J., and Gopel, W. (1999) *Anal. Chem.* 71, 2512.

Chapter 2
Selectivity

2.1 Introduction

In Chapter 1, we began the discussion of selectivity. In brief review, when an ideal selective layer is exposed to a mixture of molecules, it interacts with those for which the layer is selective and rejects the other, interfering molecules. The selective layer itself can be homogeneous or can contain specific binding sites embedded in a matrix. An outline of the thermodynamics governing the equilibrium binding was given in Section 1.1.

2.2 Equilibrium-Based Selectivity

The discussion up until now has been formulated in terms of the change of free energy. We have yet to consider the types of chemical interactions that may be involved. The interactions that are relevant are all weak interactions. They are summarized in Table 2.1, and arranged in order of decreasing strength. The covalent bond is shown only for comparison; otherwise, it has no function in the selectivity scheme.

The energies in Table 2.1 are listed as enthalpies (ΔH), but the driving forces in the chemical species/sensor interactions are really the changes of free energy (ΔG), which include the change of entropy (ΔS). At constant temperature, the two are related by (2.1).

$$\Delta G = \Delta H - T\Delta S \tag{2.1}$$

The higher the entropy change, the more negative the free energy change is and the more stable the system. Entropy is a measure of the randomness of the system (Appendix A).

There are two aspects of Table 2.1 that we need to pay attention to, in addition to the interaction energies. First is the effect of the dielectric constant. Because the

Table 2.1 Approximate interaction energies. D is dielectric constant; α is polarizability; μ is dipole moment; r is distance; z is charge; Θ is angle and I ionization energy (adapted from van Holde, 1985)

Type of Interaction	Distance relationship (nm)	Order of magnitude (kJ/mol)
Covalent bond	0.08–0.2	50–200
Hydrogen bond	0.1–0.3	20–150
Donor–acceptor	0.1–0.3	50–150
Ion–ion	$E \approx \frac{z_1 z_2}{Dr}$	90
Ion–dipole	$E \approx \frac{z_1 \mu_2 \cos\Theta}{Dr^2}$	15
Dipole–dipole (stationary)	$E \approx \frac{\mu_1 \mu_2}{Dr^3}$	± 2
Dipole-induced dipole	$E \approx \frac{z_1 \alpha_2}{Dr^4}$	2
Dispersion	$E \approx \frac{\alpha_1 \alpha_2}{r^6} \frac{I_1 I_2}{I_1 + I_2}$	2–4
CH_3–CH_3	<0.1	(1.2)[a]
$\Phi - \Phi$	<0.1	(5)[a]

[a]Means per mole unit coupling

nature of all these interactions is electrostatic, increasing the dielectric constant (D) makes most of the interactions weaker. Second is the dependence of the interaction on distance, that is, the term (r). The higher the power of r, the faster the interaction energy decays with distance between the molecule and its binding site. This is why weak interactions are also called *short-range interactions*. There is no distance dependence for donor–acceptor complexes, hydrogen bonds, or hydrophobic bonds, in which the bond length is fixed and typically on the order of tenths of nanometers. The significance of the short-range distance dependence is in the shape recognition, which is the most important reason for the high selectivity inherent in the so-called "lock-and-key" biological interactions. There, as we show, the geometry of the interaction plays a dominant role.

2.2.1 Shape Recognition

Living organisms are the ultimate sensing machines. In order to survive, an organism must accomplish three missions: to metabolize, to reproduce, and to process information. The latter means both information acquisition and processing. There are millions of chemical species in the environment, so the *selectivity* is of primary importance; the acquisition of "false signals" and/or their wrong interpretation could be anything from humorous to disastrous. The biological strategy to avoid such problems is to involve shape recognition, in other words, stereospecificity.

So, let us look at the geometry of the binding site and of the analyte in Fig. 2.1. We are interested in examining their geometrical fit, meaning: "shape recognition." Within the system shown in Fig. 2.1, they are clearly in a more random configuration if they are not associated, because they have more degrees of freedom and therefore
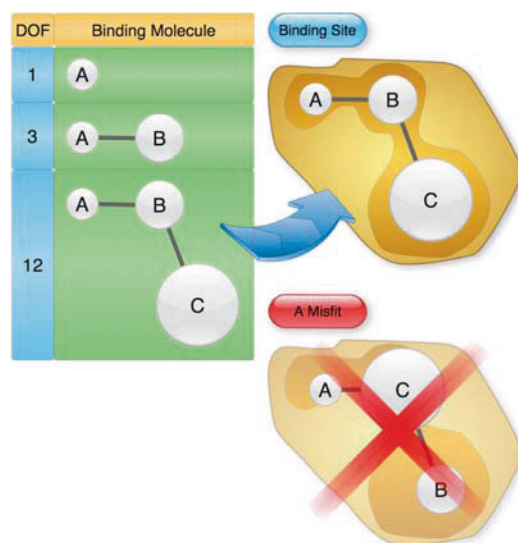
Fig. 2.1 Shape recognition by the binding site increases with the geometrical complexity of the site. The number DOF represents the degrees of freedom that must be satisfied for a perfect fit to occur

higher entropy. In other words, on the entropy consideration alone they would be more stable if they were disassociated.

The only way they would be used in an active sensor is if the enthalpy decrease due to their association were high enough to compensate for the value of the term $T\Delta S$ in (2.1). In this case, the "shape recognition" strategy will work only if there are negative-enthalpy-producing binding sites (exothermic) present inside the binding site **S**. Nature has, indeed, used this strategy in using hydrogen bonds, ionic bonds, charge-transfer complexes, and so on, inside the binding sites of hormone receptors and antibodies, among others. However, as can be seen from Table 2.1, most of the interactions fall off rapidly with the distance between the interacting entities (with the exception of the Coulombic ion–ion bond), and with the increase of the dielectric constant. The enthalpic interactions are common to many classes of compounds and as such would not be an optimum strategy for achieving high selectivity. So, is there a way out?

Let us consider the most regular object, a sphere. It has only one parameter that affects its potential to be selectively recognized: its size. There is only one "arrangement" that corresponds to a selective fit of the sphere to the binding site. Increasing the number of spheres in the arrangement can be conceptualized as increasing the number of atoms in the target molecule, which must correspond to an equal number and arrangement of spaces in the binding site if a selective fit is to be possible. Thus, if we now take two spheres connected with a solid rod (like a "dumbbell") the number of possible "fits" (indicated in Fig. 2.1 by the term **A**) increases to three: two for the sizes of the balls and one for their distance. We take one more step in this progression and interconnect three balls of different size with

two links of different length. For this "molecule" the number of requirements for a perfect fit increases to 12: three for the diameters of the balls, two for the lengths of the links, one for the angle, and then double that for the configuration 1-2-3 or 1-3-2.

It is important to realize that only one configuration will fit the binding site. Therefore, it is generally true that, as the geometrical complexity increases, the number of possible misfits decreases and therefore the selectivity increases. Moreover, the "size of the ball" can be also interpreted as their ability to interact. Thus, even if the molecule fits the binding site geometrically, it must also match the type of enthalpic interaction belonging to that part of the binding site. For example, if Region 2 on the molecule is hydrophobic then it must be matched by the hydrophobic region in the binding site. Any other interaction will result in $\Delta G > 0$, which means repulsion. Therefore, a molecule approaching the binding site can be either attracted ($\Delta G < 0$) or repulsed ($\Delta G > 0$). So, in effect, a "mismatch" may actually have a negative contribution to the binding, and thus further enhance the overall recognition.

As we see from (2.1), the increase of entropy favors binding. This is the essence of the probably most important bond in biological systems, the hydrophobic bond. Figure 2.2a depicts the situation as it exists in vacuum, which is obviously not
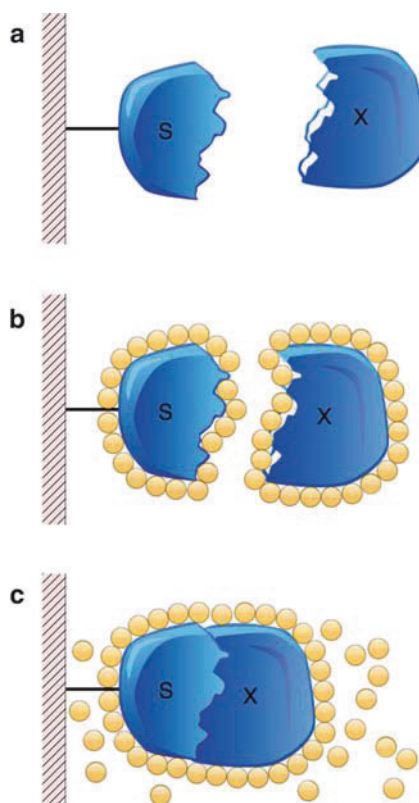
Fig. 2.2 Entropic nature of the "hydrophobic" bond

a "typical" biological situation. Biological sensory systems work in an aqueous environment. This means that water must be included in the thermodynamic considerations. Referring back to Fig. 2.2, you can see the water molecules (gray spheres) interact (hydration) with both **S** and **X**. This is because the hydrogen bond, ion–dipole, and dipole–dipole interactions exhibited by H_2O have enough energy to cause such association. When **S** and **X** are separated (Fig. 2.2b), there are arbitrarily selected two free (random) molecules of water present in the system.

For a unique description of the system we have to specify the coordinates of all the components. If we place the origin of the coordinate system in the binding site **S**, then the position of the remaining particles is uniquely described by $3n = 9$ coordinates. Those are their degrees of freedom. On the other hand, when the complex **SX** is formed (Fig. 2.2c), 14 more water molecules will have been liberated from the cleft. This increases the number of the particles to 15 and the number of degrees of freedom to $15 \times 3 = 45$ (the numbers correspond, of course, to Fig. 2.2, but are otherwise arbitrary).This increase of degrees of freedom results in the increase of the entropy (and a corresponding reduction in the free energy) of the whole system. Thus, the binding free energy is driven by the entropy of hydration. Its equivalent enthalpy value/interaction is included in Table 2.1 for comparison.

For the purpose of this discussion, it is most important to realize that the hydrophobic bond will contribute to the free energy of interaction only if the two molecules geometrically fit and eliminate some hydration water from the binding cleft and its vicinity. It is this condition that primarily accounts for the high selectivity found in the immunochemical reactions, biological receptor binding, enzyme/substrate recognition, and so on. It is also combined with enthalpy-driven binding, which can again act only at a relatively short range.

The equilibrium constants of some biological recognition reactions have a stronger dependence on temperature than others, implying that the relative contribution to free energy change from, for example, the hydrogen bond (enthalpy) and the hydrophobic bond varies (Absolom and van Oss, 1986). The important corollary for the design of so-called biosensors is that it would be difficult to employ a water-based biological selective system (e.g., most enzymes or antibodies) for water-free applications.

2.2.2 Bioselectivity

Specific binding sites of biological origin define a special class of chemical sensors: biosensors. Due to their exceptionally high selectivity, these bioligands have been the subject of intense interest among sensor scientists and engineers. They are classified in Table 2.2.

Unfortunately, high selectivity implies strong binding energies, often in excess of $100 \, kJ \, mol^{-1}$. It then becomes a case of "too much of a good thing." Once again, we must invoke the difference between the sensor and sensing system (or assay). The latter requires some intervention in order to disassociate the strong complex formed

Table 2.2 Bioligands used in biosensors and bioassays

Bioligand	Use
Antibodies/antigens	Immunoassays
Oligonucleotides	DNA (RNA) bioassays
Aptamers	Bioassays
Enzymes	Enzyme sensors
Receptors	Bioassays
Cells and tissues	Bioassays

between the biological ligand and the substrate. Such a step places them outside the definition of the chemical sensors and outside the scope of this book. However, they are briefly discussed here for several reasons.

First, it may help to clarify the misuse of the term "biosensor." Perhaps more important, there is a real hope that these bioligands could be used in genuine biosensors if their high binding constant could be lowered to the level where they could be used in an equilibrium-binding regime. Such de-tuning of binding is possible, in principle, if the binding region in the biomolecule is covalently modified. Another possibility is to operate them under such conditions that the binding interaction is weakened to the point that equilibrium condition is established. This can be done by modifying the reaction medium, that is, increasing ionic strength and lowering the pH (immunochemical interactions), addition of "hydrogen bond breakers" (DNA, RNA binding), elevating the temperature (DNA, RNA binding), or use of mixed organic/aqueous solvents (immunochemical and receptor binding). Such a change of operating conditions may, however, impose unacceptable constraints on the operation of the biosensor. Finally, their very successful use in various bioassays is another reason for having them briefly discussed. This assessment applies to all bioligands except enzymes. They stand apart and are discussed in the section on kinetic selectivity (Section 2.3).

2.2.2.1 Immunochemical Selectivity

If one were to define the ideal building blocks for the construction of a selective layer, antibodies would have to be considered very seriously. Their selectivity is based on stereospecificity of the binding site for the antigenic determinant (antigen, hapten, epitope). Their production is relatively inexpensive and universal, which means that an antibody for any antigen, regardless of its shape or chemical nature, can be produced by the same general procedure. The only limitation seems to be that of the size: antigens of a molecular weight of less than 2,000 D normally do not induce an immunochemical response in B lymphocytes that produce them. In order to obtain antibodies for low molecular weight antigens (haptens) it is necessary to link the latter to a high molecular weight polymeric carrier (e.g., bovine serum albumin [BSA], polyethylene glycol, etc.).
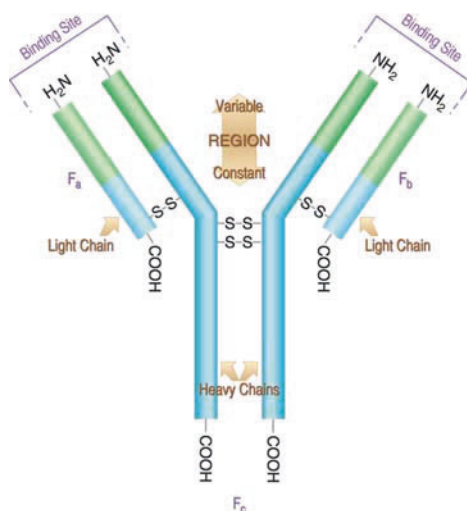
Fig. 2.3 Schematic of antibody immunoglobulin G (IgG). It is a bivalent bioligand with binding sites formed by light-chain fragments F_{ab}

Antibodies belong to the group of serum proteins called immunoglobulins (Janeway et al., 2004). Their molecular weight ranges from 140 to 970 kD. The number of antigens that can be bound to one antibody determines their valency, which is typically 2 but can be as high as 10 for immunoglobulin M (IgM). Their primary function is to disable foreign (high molecular weight) immunogens, be it proteins, nucleic acids, viruses, and so on that may invade and endanger the organism. In that respect, they can be looked at as highly specific complexing agents, which are one of the key factors in the defense mechanism. The most common antibody is immunoglobulin G (IgG) which has a molecular weight of 146,000 D and valency of 2 (Fig. 2.3). Traditionally, it has been produced by multiple sensitization (inoculation) of an experimental animal (e.g., rabbit, goat, dog, etc.) to a suitable antigen. When the natural immune response triggers the production of IgG, which is targeted against the immunogen, the specific antibody can be isolated from the animal antiserum and purified to the desired level. Antibodies produced by this procedure are called polyclonal. A more efficient procedure involves fusion of the sensitized B lymphocytes with myeloma cells (malignant cancer cells), which are then implanted in the animal. They produce large amounts of specific monoclonal antibodies, which are again harvested and purified.

Although the active site on the antibody is fundamentally highly specific to the given antigen, any preparation of antibodies (either poly- or monoclonal) is heterogeneous. This heterogeneity is far greater in polyclonal antibodies than in their monoclonal counterparts. The average ability to complex antigen is called the avidity of the preparation, and the binding equilibrium between an antibody (Ab) and an antigen (Ag) is referred to as affinity.

$$Ab + 2Ag \underset{k_r}{\overset{k_f}{\leftrightarrow}} [Ab * Ag_2] \tag{2.2}$$

The binding constant K is again defined as the ratio of the forward (k_f) and reverse (k_r) rate constants (2.2). It ranges from $K < 10^4 \, L \, M^{-1}$ for weak binding of antigens of MW $< 2,500$ D to $K > 10^9 \, L \, M^{-1}$ for antigens of MW > 6–$8 \, 10^6$ D, which are very strongly bound.

Under physiological conditions, an equilibrium constant in the range of $K = 10^5$ to $10^9 \, L \, mol^{-1}$ corresponds to a ΔG^0 range of -25 to $-50 \, kJ \, mol^{-1}$. The forward rate of the immunochemical reaction is invariably very high (diffusion-limited). This is consistent with the strategy of the biological defense mechanism, where the inactivation of a "potentially harmful" antigen must be done with maximum speed, but the recognition of the truly harmful (or innocuous) constituent can be done much more slowly. This means that the dissociation rate constants vary over 7 decades from $10^{-4} \, s^{-1}$ to $10^3 \, s^{-1}$ and determine the overall high affinity of the hapten or antigen to the antibody.

The nature of the Ab–Ag bond is of critical importance for analytical purposes. The most prevalent bonds are considered to be Coulombic and van der Waals interactions. The role of water in the overall binding is also critically important. First of all, it is the prerequisite in the formation of the hydrophobic bond. However, expulsion of water from the binding site, which takes place during binding, decreases the local dielectric constant and increases the strength of the Coulombic and van der Waals bonds (Table 2.1) in that region. Thus, the binding is cooperative. The close stereospecific fit is, of course, necessary. There is no covalent bonding involved in any immunochemical reactions.

The binding equilibrium expressed as shown above (2.2) is actually a gross oversimplification of the situation. The heterogeneity of the binding sites and multiple valency of individual antibodies lead to formation of secondary bonds that contribute to hysteresis or "ripening" of the antibody–antigen complex. Its ultimate form is the polymerization of a primary complex, which happens when the antigen is also polyvalent. Formation of the polymer (precipitin reaction) renders such a reaction virtually irreversible.

The secondary bonds, which may be formed much more slowly than the primary bonds, actually contribute more to the overall affinity. For example, the primary (Coulombic) bond between bovine serum albumin (BSA) and anti-BSA IgG is $3.3 \, kcal \, M^{-1}$ whereas the secondary bond (van der Waals) is $28 \, kJ$, for a total $\Delta H = 42 \, kJ$. Because the formation of the secondary bond is much slower, it is easier to prevent formation of the strong complex rather than to try to dissociate it. This is one reason why the competitive immunoassays yield results that correlate with the equilibrium-binding constants, but any such direct-binding assays have to rely on the measurement of the initial rate of binding.

In order to assess the utility of the immunochemical reaction for chemical sensing, we need to examine the effects of the experimental conditions on the primary association reaction. The effect of temperature is not particularly distinct for most reactions and cannot be generalized. This is due to the fact that the relative

contribution of the hydrophobic (entropic) bond and other (enthalpic) bonds is different. The equilibrium is largely insensitive to pH (between 6.5 and 8.5) and normal ionic strength. However, lowering the pH below 2 and increasing the ionic strength above 1 M weakens the Ab^*Ag complex to the point that it can be dissociated. The presence of organic solvents begins to play a role only when the hydrophobic bonds become affected. Obviously, the presence of water in the binding process is mandatory.

The dissociation of the "aged" immunochemical complex can be achieved, but sometimes only under denaturing conditions. The techniques that have been used mainly for preparative purposes include use of chaotropic salts (e.g., KCNS, tetraethylammonium chloride, guanidine hydrochloride, etc.). These salts compete for "available" water. Ionic (sodium dodecyl sulfate) and nonionic (polyethyleneglycol) detergents as well as solvents (ethanol, dimethyl sulfoxide) play a similar role. In general, these agents tend to disrupt the hydrophobic bond and have to be applied in relatively high ($\sim$M) concentrations. Disruption of hydrogen bonds (which leads to denaturation in most cases) is done by addition of 6–8 M urea and/or by lowering the pH to approximately 2. It has been possible to dissociate the charged antigen by placing the complex in a high electric field. This is so-called electrophoretic dissociation, which works only in solutions of very low ionic strength. Finally, it is sometimes possible to use low molecular weight haptens, particularly in the early stages of complex formation. In that sense the dissociation with the help of hapten can be regarded as a competitive binding.

Are antigens really the ideal binding sites for sensor applications? The answer to this question is not straightforward. First of all, the high (and time-variable) value of the affinity binding constant of particularly polyclonal antibodies makes the interaction virtually irreversible. We have to combine this with the fact that the IgG molecule is large, and that the area occupied at the sensor surface by one active site is also very large. This means that the packing density is very low and the dynamic range is narrow. This situation can be somewhat improved by isolating the binding sites from the rest of the molecule. Furthermore, we have to eliminate the multivalency in order to prevent polymerization. Again, this can be achieved at the level of a single F_{ab} fragment, as shown in Fig. 2.3. The cost of preparing single F_{ab} fragments is, however, much higher. A good strategy for improving the dynamic range would be to increase the number of binding sites by immobilizing them in a layer, rather than on the surface, and to incorporate sites with widely different values of the binding constant. The formation of the secondary bonds should be prevented. There is no simple recipe for doing this, except a mode of operation resembling competitive binding.

In contrast with sensors, sensing systems are ideal for exploitation of immunochemical selectivity. This accounts for various highly sensitive and successful competitive immunoassays. Incorporation of the manipulative step(s) opens the door for regeneration of the antibody or even for operation under virtually reversible conditions.

2.2.2.2 Nucleotide-Based Selectivity

The selectivity of DNA (and RNA) interaction is probably the highest of all biological recognition sites. It is unique in that it relies exclusively on highly stereospecific hydrogen bonding between base pairs: adenine–thymine and cytosine–guanine (Fig. 2.4). The enthalpic value of one base pair formation is $\Delta H = 20.1\,\text{kJ mol}^{-1}$ for the A–T and $\Delta H = 57.5\,\text{kJ mol}^{-1}$ for the G–C. Because the interaction enthalpies are additive, the overall DNA fragment increases with the number of base pairs, reaching the "reversible" threshold even for a dimer. The sensing dilemma is obvious: the interaction is again too strong. The sensing reversibility can be achieved, in principle, by operating the sensor near the "melting temperature" of the duplex, but the melting point depends on the number and type of base pairs and is not generally known a priori.

It is necessary to invoke the meaning of selectivity at this point. If the DNA fragment contains a "mismatch," it must be considered to be an interferant, or impurity, in the sensing context. However, one such mismatch may lower the interaction
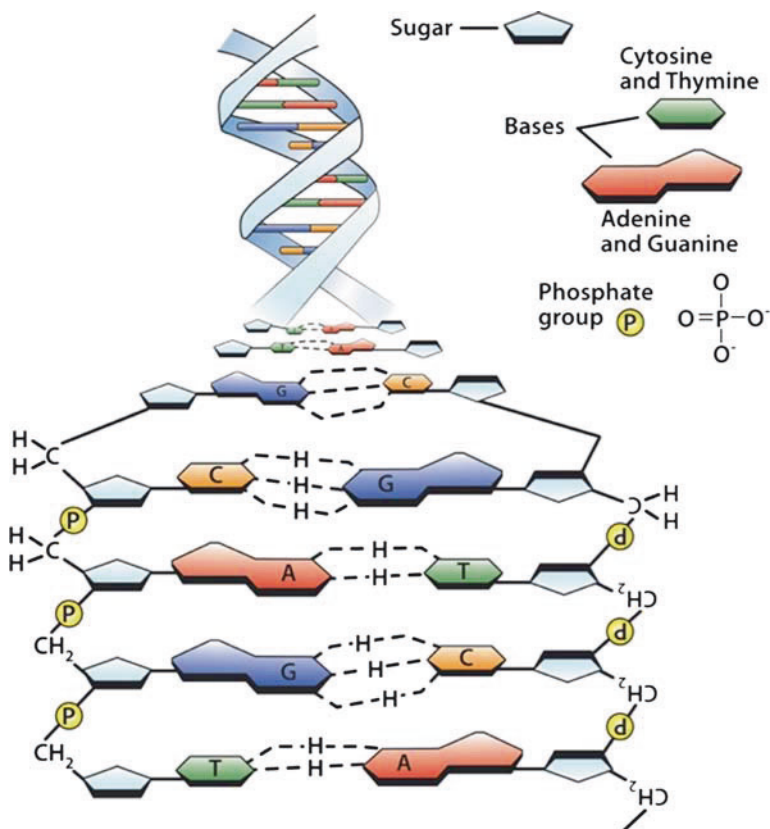
Fig. 2.4 Hydrogen bonds in DNA base pairings

energy by such a small amount that for all purposes the binding and sensing still take place. If the transduction mechanism cannot distinguish such an event, then the advantage of high selectivity is lost. Thus, paradoxically, the high selectivity of the elementary sensing interaction becomes self-defeating. Moreover, the position of the mismatch in the DNA duplex also plays a role. The hybridization process is sequential in nature. This means that the association process starts at one end of the single strand (ssDNA) and progresses down the chain. This is known as the "kiss-and-zip" mechanism. A single mismatch is always skipped and does not play a major role in the overall result, except for a slightly lower overall binding energy and melting temperature.

2.2.2.3 Aptamers

These are "artificial/natural" oligonucleotides (DNA or RNA), in which the principle of the biological "lock-and-key" recognition is preserved (Fig. 2.5a).

They are capable of binding small molecules in the range of 100–10,000 D. They have been designed for assays of drugs, small proteins, and other small molecules (Tombelli et al., 2005). Their affinity is comparable to, or higher than, corresponding monoclonal antibodies. It is due to the unique folding ability of RNA and of single-stranded DNA. They are prepared by an entirely in vitro procedure called
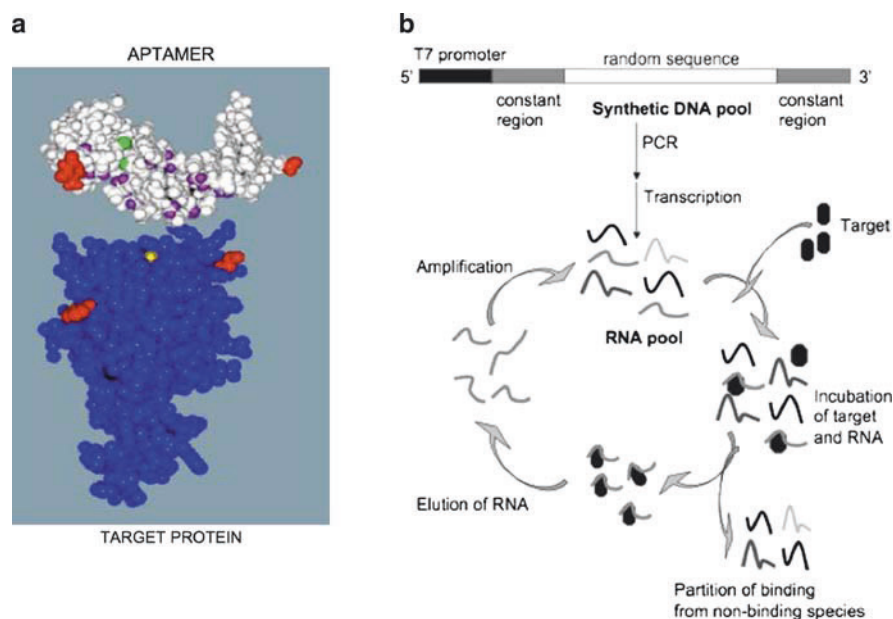
Fig. 2.5 (**a**) An aptamer-target protein interaction (adapted from Tuerk and Gold 1990, p. 505). (**b**) Principle of the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) process (adapted from Tombelli et al., 2005, p. 2424)

the SELEX (Systematic Evolution of Ligands by Exponential Enrichment) process (Tuerk and Gold, 1990; Fig. 2.5b). It is obvious that they suffer from the same problem as other bioligands with a high binding constant, that is, a virtual irreversibility. There are a number of applications in bioassays (Tombelli et al., 2005) where they show better performance (namely long-term stability) than antibodies. What makes them potentially interesting for reversible sensing applications is the opportunity to manipulate, specifically to *decrease*, the binding constant to the point that they would operate in the reversible regime.

2.2.3 Imprinted Polymers

Another way to realize the shape recognition ability is through the process known as "molecular imprinting" (Diaz-Garcia and Badia, 2004; Haupt, 2004). The process is depicted in Fig. 2.6.

In this process, a template molecule creates a "footprint" in the polymerizing matrix. After its removal from the polymerized material this footprint becomes a shape-recognizing specific binding site for the same molecule. The idea of molecular imprinting is quite old, dating to the mid-1950s when Linus Pauling reported selective sorbents from silica gels. Imprinted polymers came later and have been successful as stationary phases in chromatographic separations, particularly of chiral isomers. In applications as direct sensing materials, however, their success has been quite limited.
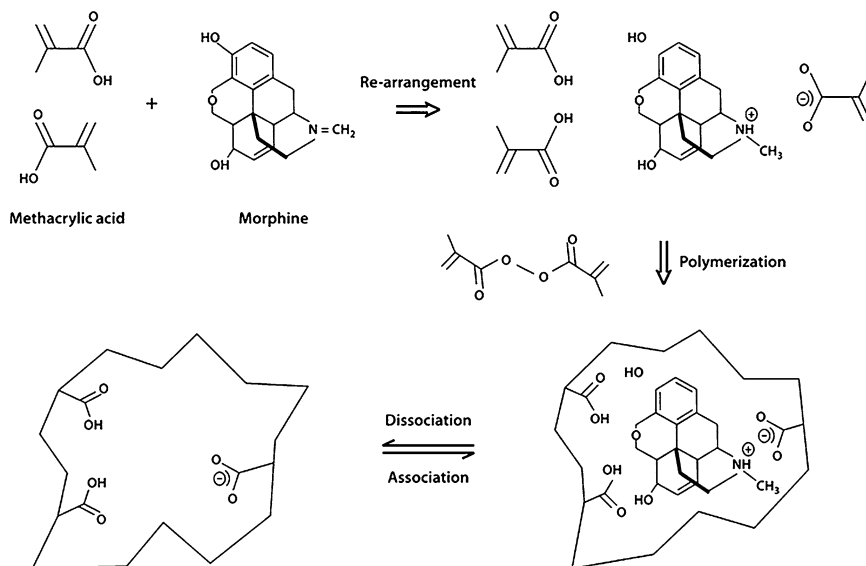
Fig. 2.6 Principle of molecular imprinting (adapted from Diaz-Garcia et al., 2004)
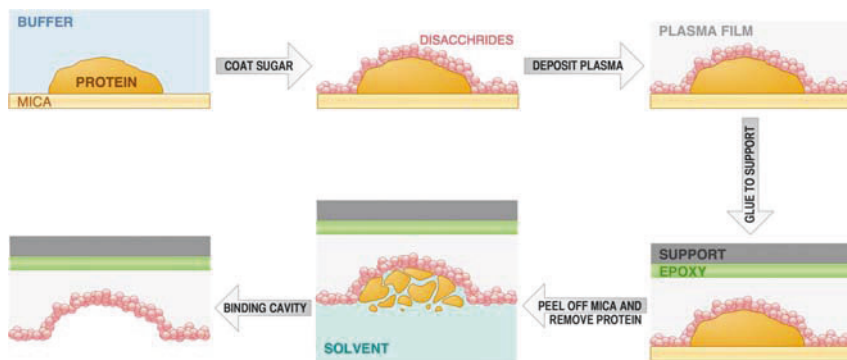
Fig. 2.7 Creation of an "imprinted" surface

There are two processes by which the bulk imprinted polymers are formed: covalent imprinting and noncovalent imprinting. In the former, the template molecule is first covalently functionalized with the monomer, and then copolymerized with the pure monomer. After that the covalent bond is broken and the template molecule is removed by extraction. In order to facilitate the extraction step, a so-called porogenic solvent is used. It effectively swells the polymer matrix.

In the noncovalent approach, the monomer is self-assembled around the templating molecule and then again copolymerized with the additional monomer. The template is then removed by using a porogenic solvent.

It is also possible to prepare molecularly imprinted surfaces (Huaiqiu Shi et al., 1999). The process is depicted in Fig. 2.7. First the template (in this case protein) is deposited on a mica surface, which is atomically smooth. Next, the surface and the adsorbed protein are coated with water-soluble disaccharide (sugar). After that, a fluoropolymer is deposited by plasma polymerization of C_3F_6. A mechanical support is then added by attaching a glass coverslip with epoxy. Finally, the mica support is peeled off and the sugar coating and protein molecules are washed away, exposing the "footprint pit" where the protein was. The thus-prepared surface shows up to a tenfold preferential enhancement of adsorption of the template protein molecule as compared to other nonimprinted proteins.

There are several reasons why imprinted polymers do not match the affinities of natural stereospecific binding sites. First of all, the shape alone (i.e., entropic contribution) is not sufficient. The specific short-range interactions that exist in the natural binding sites generally do not exist in the imprints. Second, it is implicit in solution polymerization that the templating molecules and the monomers are solvated and that the solvation shell contributes to the overall shape and size of the template in a significant way. When the imprinted material is used in a solid–gas interaction, the fit of the molecule without the solvent is poor. Another reason is that a "tight fit" implies that the template is locked in the bulk of the polymer matrix and cannot be extracted, even with the aid of the porogenic solvent. The fact that it is leached out and that the template molecule can exchange between the imprint and the sample is because the fit is not as good as it has been expected. Therefore,

the binding reversibility and the stereoscopic specificity are two conflicting requirements. This problem does not exist on the imprinted surface, where the binding pit is freely accessible. Only properly designed baseline experiments can truly assess the viability of the imprinting approach.

2.2.3.1 Solubility in Organic Materials

The main difference between the materials described in this section and the previous ones is that these materials are homogeneous. In other words, there are no discrete binding sites and the interaction between the analyte and the selective layer is governed by the Gibbs equation (see (2.1)). The solid phase is treated as a solid solvent to which the analyte partitions from the sample. The solid phase can be amorphous or polycrystalline or even a gel and the sample can be liquid or gas. An immediate analogy comes to mind: a gas chromatographic (GC) experiment. There, the sample partitions between the mobile gas or liquid phase and the stationary, solid, or semiliquid phase. Indeed, such an analogy leads to one of the most successful empirical relationships, the Linear Solvation Energy Relationship (LSER) that has been used in the design of selective layers, particularly for gas sensing. The partitioning process is described by (2.3).

$$\Delta G_S^0 = -RT \ln K_S = c + rR_2 + s\pi_2 + a\Sigma\alpha_2^H + b\Sigma\beta_2^H + l\log(^{16}L) \qquad (2.3)$$

The lower-case coefficients are related to the sorbent material and the capital and the Greek letters describe the gas. Hundreds of these values have been compiled and tabulated from the GC data (Abraham, 1993).

The terms on the right-hand side of (2.3) have the meaning of individual contributions to the Gibbs free energy change according to specific interactions, more or less matching those given in Table 2.1. The second term, rR_2, is the polarizability, which describes the interactions involving induced dipoles. The term $s\pi_2$ is the polarity, describing ion–dipole and dipole–dipole interactions. The terms $a\Sigma\alpha_2^H$ and $b\Sigma\beta_2^H$ relate to hydrogen bonding at acidic (a) and basic (b) sites, respectively. Finally, the last term ($l\log(^{16}L)$) is related to the dispersion of van der Waals interactions. The superscript 16 indicates the carbon-16 alkyl chain against which the dispersion has been referenced.

The usefulness of the LSER approach hinges on the similarity of the partitioning coefficients obtained from the sensing experiments (K_S) and the gas chromatographic experiments (K_{GC}). In other words, it is assumed that the relationship $K_S \approx K_{GC}$ holds. This is how LSER is used for evaluation of a new sensing material. First, the coefficient K_{GC} is obtained from the tabulated database or experimentally. Second, the multiple linear regression technique (see Chapter 10) is used to obtain the best fit for the sensor test data, and the individual coefficients in (2.3) are evaluated. This approach has been used successfully in evaluation of multiple materials for gas sensors (Abraham et al., 1995; Grate et al., 1996).

The coefficient c in (2.3) is a fitting parameter; it does not have an assigned physical meaning, but may account for the difference between the static (K_S) and dynamic (K_{GC}) nature of the two experiments. It has been found that the GC partitioning coefficients are consistently lower by a factor ~ 4, than those obtained from the mass sensor measurements with QCM and SAW sensors (Chapter 4; Hierlemann et al., 2001). This discrepancy may have its origin in the different nature of the two experiments. In the chromatographic experiment, the gas molecules at the front of the advancing zone encounter "pristine" sorbent material. This is particularly important when a mixture of analytes is evaluated. Second, the GC measurement is dynamic and not done under the conditions of fully developed equilibrium. Nevertheless, in spite of this discrepancy, the predictive properties of LSER have been exceptionally successful in the design of new sensing materials. The main domain of LSER application has been in the design of selective layers for various types of mass sensors.

One type of interaction that is not covered in the LSER equation (2.3) is formation of the charge-transfer complex, which can also increase the solubility of the gas on the selective matrix. Partial transfer of charge in electron donor–acceptor interactions is a common notion in organic chemistry (Reichardt, 1988). The bond that is formed is a dipole whose dipole moment depends on the fraction of transferred charge (δ), and on the separation distance. When this interaction takes place between two molecules, the positive end of the dipole (δ^+) is located at the donor molecule and the negative (δ^-) at the acceptor molecule. The amount of transferred charge depends on the electron affinities of the participating molecules. The notion of electron affinity applies also to electronically conducting solid phases where it is related to the position of the Fermi level and the value of work function (Appendix D). If the material has a high value of work function it will act as an electron acceptor and vice versa. Therefore, molecules of gas that have low electron affinity (i.e., low ionization potential) will partially transfer electrons to the conduction energy band of the material and become associated with the matrix. From the material's point of view, this guest–host interaction represents a form of doping that changes the electronic properties of the material, namely its conductivity and work function. From the guest molecule viewpoint, it increases its solubility in the matrix. The geometrical arrangement of this association is highly specific to the material.

It is important to realize the crucial difference between charge-transfer doping and ionization doping. In charge-transfer doping, it is the electrically neutral molecule that interacts with the solid matrix. Such an interaction is typical for gases and can be exploited in gas sensors. On the other hand, in the ionization doping process, the electron is completely exchanged between the guest molecule and the matrix, leaving the usually immobile donor cation (in n-doping) or immobile acceptor anion (in p-doping). In chemists' language, the charge-transfer doping process constitutes Lewis acid–base chemistry whereas the ionization doping is characteristic of the ionization or oxidation-reduction (redox) process. This distinction is critically important for chemical sensing. The ionization doping is the key mechanism of ion-selective electrodes (Chapter 6) where the ion selectively partitions into the organic phase, called the ion-selective membrane. On the other hand,

charge-transfer doping is the key mechanism in work function sensors (Chapter 7; Janata and Josowicz, 1998). The transduction principles that apply to these two interactions are substantially different. The two mechanisms can be represented as follows.

Charge-transfer doping:

$$OS_{\text{solid}} + G_{\text{sample}} \Leftrightarrow \left[OS^{\pm\delta} * G^{\mp\delta} \right]_{\text{solid}} \tag{2.4a}$$

Ionization doping:

$$OS_{\text{solid}} + I^{\pm}_{\text{liquid}} \Leftrightarrow [OS \cdot I]^{\pm}_{\text{solid}} \tag{2.4b}$$

The asterisk in (2.4a) indicates that the guest molecule G is associated with the organic semiconductor (OS) in some intimate, dipolar geometrical arrangement dictated by the partially exchanged charge δ. Because the guest molecule is electrically neutral, the sample from which this molecule can be partitioned can be either gas or liquid. However, partitioning of ions applies to partitioning between OS and a liquid sample. The organic materials that fall into this category of selective materials are all organic semiconductors, namely conducting polymers, redox polymers, and van der Waals organic solids (Janata and Josowicz, 1998). Because there are many more electrically neutral gases to be detected than ions, the charge-transfer doping is potentially the more prevalent type of interaction.

It was mentioned at the beginning of this section that the distinguishing feature of materials based on solubility-based selectivity is that they are homogeneous. However, this statement should not be taken too literally, because enhancement of selectivity can be achieved by incorporating specific binding sites into these matrices. The choice of correct transduction mechanism then depends on the type of analyte/selective layer interaction. Generally speaking, detection of mass change will work in all modes of solubility-based selective materials, as long as the mechanical properties of such layers are not affected (Hierlemann et al., 2001; Topart and Josowicz, 1992).

2.2.3.2 Solubility in Inorganic Materials

Although the general principle of partitioning equilibrium remains the same, there are additional mechanisms, and the underlying physical principles governing inorganic materials are different. An important, albeit somewhat unique, example is the solubility of hydrogen in palladium metal in which the charge-transfer mechanism again applies. Molecular hydrogen first dissociates into atomic hydrogen (Ekedahl et al., 1998), which then diffuses into the Pd bulk, forming bulk palladium hydride PdH_x (Fig. 2.8).

Thus, at the surface

$$H_{2,\text{ads}} \Leftrightarrow 2H_{\text{surf}} \tag{2.5}$$

This is followed by

$$Pd_{\text{surf}} + x H_{\text{surf}} \Leftrightarrow PdH_{\text{bulk}} \tag{2.6}$$
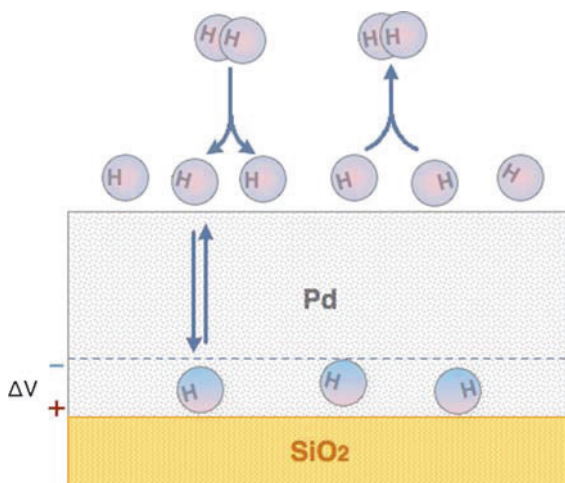
Fig. 2.8 Palladium/hydrogen interaction

In the absence of oxygen, the reaction depicted by (2.5) is reversible. However, when oxygen is present, a competing oxidation takes place at the Pd surface, making the overall reaction irreversible.

$$4H_{surf} + O_2 \Rightarrow 2H_2O \uparrow \tag{2.7}$$

Hydrogen is a very important species and various hydrogen sensors based on reactions (2.5) through (2.7) have been commercialized. They are discussed in more detail later.

An entirely different selectivity principle known as phase equilibrium comes into play in high-temperature ionic conductors. Many important gases dissolve in ionic solids at elevated temperatures. However, the solubility is rather sharply defined for the gas and the solid by the lattice parameters and the size of the gas molecule. The best example is the solubility of oxygen in zirconium dioxide. When ZrO_2 is doped with yttrium ions, it exhibits a high mobility for the O^- anion. The solubility and anion mobility then become the basis for several electrochemical gas sensors, using "yttria-stabilized zirconia" (YSZ).

There are several factors that make solid-state ionic conductors attractive for chemical sensing purposes. One is the aforementioned selectivity stemming from the narrowly defined solubility. Second is the fact that these materials are intensively investigated as building blocks of fuel cells and the knowledge database (useful also for sensors) is rapidly expanding. Third is their operating temperature, which is typically well above the boiling point of water, eliminating this ubiquitous interference so often found in room-temperature sensors. (The reverse side of this coin is that the high temperature requirement somewhat limits their applications.) Finally, the diffusion is faster at higher temperatures, resulting in faster response time.

2.3 Kinetic Selectivity

This form of selectivity applies to sensors that operate in the steady-state regime. The prime examples are thermal and amperometric sensors. It is somewhat limited for potentiometric sensors and it is least suitable for mass sensors. The minimum necessary kinetic background information can be found in Appendix B.

Consider a mixture of species, X_1, X_2, ..., X_n, which are undergoing common chemical transformation to products P, but with different reaction rates.

$$X_1 \xrightarrow{k_1} P_1 \tag{2.8a}$$

$$X_2 \xrightarrow{k_2} P_2 \tag{2.8b}$$

$$X_n \xrightarrow{k_n} P_n \tag{2.8c}$$

Let us assume that a catalyst can selectively increase the rate of conversion of the analyte X_X.

$$X_X \xrightarrow[\text{cat.}]{k_X} P_X \tag{2.9}$$

Therefore,

$$k_X >> k_1, k_2, \ldots, k_n \tag{2.10}$$

Such a situation may arise, for example, in the combustion sensor in which the species of interest is methane and the other combustibles are different higher molecular weight hydrocarbons. The catalyst, in this case, can be Pt and the preferentially catalyzed reaction is as follows.

$$2CH_4 + 3O_2 \xrightarrow[Pt]{\Delta H} 2CO_2 + H_2O \tag{2.11}$$
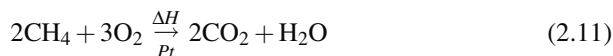
This is the reaction taking place at the surface of the thermal sensor, the pellistor, discussed in Chapter 3. An example of a biocatalyst is the enzyme glucose oxidase (GOD) which highly selectively promotes oxidation of D-glucose to gluconic acid.

$$C_6H_{12}O_6 \xrightarrow{\text{GOD}} C_6H_{12}O_7^- + H^+ \tag{2.12}$$
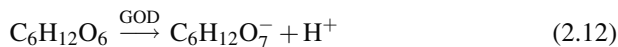
This reaction is used throughout this book because it is the most common biocatalyst in the biosensor literature and it is discussed in greater detail below.

2.3.1 Enzyme Kinetics

In terms of sensing applications, enzymes vastly outnumber any other type of catalysts. They are natural products in biological systems where their primary function is to control the rates of important reactions, mainly, but not exclusively, in metabolism. There are a few lipophilic enzymes, but for the most part they function

in an aqueous environment. Enzymes are the key component in the largest group of biosensors. In the following section we outline the fundamentals of enzyme kinetics. The specific differences that arise from the different transduction mechanisms of different sensors are discussed separately. Here, we focus only on the key aspects of enzymatic reactions.

Enzymes are a special kind of catalyst, proteins of MW 6,000–400,000 which are found in living matter. They have two remarkable properties: (1) they are extremely selective to the given substrate; and (2) they are extraordinarily effective in increasing the rates of reactions. Thus, they combine the recognition and amplification steps. A general, enzymatically catalyzed reaction can be described by the Michaelis–Menten mechanism, in which E is the enzyme, S is the substrate, and P is the product, formed from the intermediate complex ES.

$$S + E \underset{k_{-1}}{\overset{k_1}{\longleftrightarrow}} ES \overset{k_2}{\rightarrow} P \tag{2.13}$$
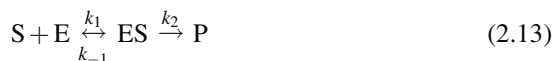
The reaction velocity (v) can be expressed as the rate of increase of the concentration of the product P.

$$v = \frac{dC_P}{dt} = k_2 C_{ES} \tag{2.14}$$

For a high value of substrate concentration, the reaction velocity reaches its maximum (saturation). Under those conditions, all the available enzyme E_T is bound in the complex with the substrate. Thus

$$v_{max} = k_2 C_{E_T} \tag{2.15}$$

This means that the maximum velocity is proportional to the concentration of the enzyme. Below saturation, the enzyme is present either in free form or complexed with the substrate.

$$C_{E_T} = C_E + C_{ES} \tag{2.16}$$

At steady state, the concentration of the ES complex is constant.

$$\frac{dC_{ES}}{dt} = k_1 C_S C_E - (k_{-1} + k_2) C_{ES} = 0 \tag{2.17}$$

The Michaelis–Menten constant (K_m) is defined as

$$K_m = \frac{k_{-1} + k_2}{k_1} = \frac{C_S C_E}{C_{ES}} \tag{2.18}$$

Substitution for C_E from (2.16) into (2.18) yields

$$K_m = \frac{C_S (C_{E_T} - C_{ES})}{C_{ES}} \tag{2.19}$$

This, when combined with (2.14) and (2.15), gives

$$K_{\mathrm{m}} = \frac{C_{\mathrm{S}}(v_{\max} - v)}{v} \tag{2.20}$$

After rearrangement, we obtain the Michaelis–Menten equation.

$$v = \frac{v_{\max} C_{\mathrm{S}}}{C_{\mathrm{S}} + K_{\mathrm{m}}} \tag{2.21}$$

It can be shown that K_{m} equals the concentration of the substrate at which the reaction velocity is one half of its maximum. The Michaelis–Menten constant is an important figure of merit for the enzyme. It is the measure of its activity. Although it describes a kinetic process, it has the physical meaning of dissociation constant, that is, a reciprocal binding constant. It means that the smaller the K_{m} is, the more strongly the substrate binds to the enzyme.

The extraordinary specificity of enzymatic catalysis is due to the shape recognition. Enzymes are proteins having a stereospecific binding site. At this site, the two reactants (in the above example, D-glucose and oxygen) are brought together in a precise and favorable orientation for the reaction to take place.

As with any other proteins, enzymes are subject to acid–base equilibria that affect their catalytic properties, that is, their K_{m} value. Each enzyme has its own characteristic pH dependence $\Re_{\mathrm{pH}}$. Thus, the general Michaelis–Menten equation, which takes into account this pH dependence of K_{m}, can be written as in (2.22).

$$v = \Re_{\mathrm{pH}} \frac{v_{\max} C_{\mathrm{S}}}{C_{\mathrm{S}} + K_{\mathrm{m}}} \tag{2.22}$$

In addition to hydrogen ions, other species can also affect the enzymatic catalytic activity. This phenomenon is called inhibition; it may be specific, nonspecific, reversible, or irreversible. The inhibition reactions can also be used for the sensing of inhibitors. The best-known example is the sensor for detection of nerve gases. These compounds inhibit the hydrolysis of the acetylcholine ester which is catalyzed by the enzyme acetylcholine esterase. Acetylcholine ester is a key component in the neurotransmission mechanism.

Enzymatic reactions combine substrate specificity with a high amplification factor. From that viewpoint they are ideal selective layers for chemical sensors. However, they are not specifically part of the information acquisition/processing scheme in nature. Their exclusive role is to lower, highly selectively, the activation energy barrier of certain reactions, thus acting as regulators. A general diagram of an enzymatically coupled chemical sensor is shown in Fig. 2.9.

The geometry shown here corresponds to a semi-infinite planar diffusion. Other geometries (e.g., radial geometries) typical for microsensors can be used. The enzyme-containing layer is usually a hydrogel, whose optimum thickness depends on the enzymatic reaction, on the operating pH, and on the activity of the enzyme (i.e., on the K_{m}). Enzymes can be used with nearly any transduction principle, that is, thermal, electrochemical, or optical sensors. They are not, however, generally suitable for mass sensors, for several reasons. The most fundamental one is the fact
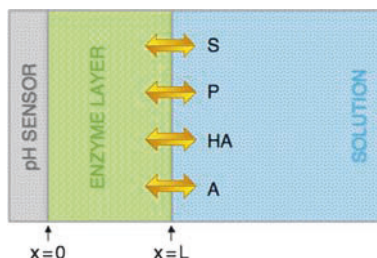
Fig. 2.9 Zero-flux-boundary enzymatic sensor

that the net mass change in a catalyzed reaction is usually small. Moreover, the mass sensors do not perform well in a gel, due to the mechanical damping.

The basic operating principle of enzyme use in sensors is simple: an enzyme is immobilized inside a permeable layer, into which the substrate(s) diffuse and from which the product(s) can effuse. Any other species that participate in the reaction, such as buffers, must also diffuse in and out of the layer (see Fig. 2.9). Because of the combined mass transport and chemical reaction, this scheme is often referred to as the diffusion–reaction mechanism.

Mathematically, this case is described by the set of second-order partial differential equations, which are usually solved numerically. The general unidirectional (in the x-coordinate) diffusion–reaction equation for any species i, is

$$\frac{\delta C_i}{\delta t} = D_i \frac{\delta^2 C_i}{\delta x^2} \pm \Re_{pH}(C_i) \qquad (2.23)$$

Here, t is time and x is the distance traveled within a layer of thickness L. D_i is the effective diffusion constant of the species i. The first term represents the mass transport, and the second is the pH-dependent "reaction term." This equation has to be written for every participating species, with the appropriate sign in front of the reaction term.

When the pH-dependent Michaelis–Menten equation (2.22) is substituted for the pH-dependent reaction term $\Re_{pH}(C_i)$, we obtain for the substrate S at any point inside the enzymatic gel layer

$$\frac{\delta C_S}{\delta t} = D_S \frac{\delta^2 C_S}{\delta x^2} - \frac{\Re_{pH} v_{max} C_S}{(C_S + K_m)} \qquad (2.24)$$

The first term on the right-hand side of (2.24) is the diffusion term, where D_S is the effective diffusion constant of the substrate. The second term is called the kinetic (reaction) term. It is necessary to normalize the variables as follows.

$$t = \frac{t^*L^2}{D_S} \tag{2.25a}$$

$$C_S = C_S^* K_m \tag{2.25b}$$

$$x = x^* L \tag{2.25c}$$

In (2.25a–c), the variables with the asterisk are dimensionless. Substitution into (2.24) yields a dimensionless diffusion–reaction mechanism equation.

$$\left(\frac{\delta C_S}{\delta t}\right)^* = D_S \left(\frac{\delta^2 C_S}{\delta x^2}\right)^* - \phi^2 \left(\frac{C_S}{1+C_S}\right)^* \tag{2.26}$$

The parameter ϕ is called the Thiele modulus.

$$\phi = \frac{L v_{max}^{1/2}}{(K_m D_S \Re_{pH})^{1/2}} \tag{2.27}$$

It contains all important design parameters, as well as the pH-dependency of the enzyme activity. It defines two operating regimes: for $\phi > 10$, the mechanism is diffusion-controlled, and for a Thiele modulus $\phi < 5$, it is reaction-controlled. In other words, it defines which of the terms on the right-hand side of (2.26) controls the rate of the conversion of the substrate. Because the terms are operating in series, the smaller of the two is dominant. For chemical sensing, diffusion control is always preferable. In order to have a high value of the Thiele modulus, we want to increase the thickness of the layer (L), decrease the effective diffusion constant of the substrate (D_S), and increase the enzyme loading (i.e., v_{max}). The value of K_m is a given for the enzyme. However, if we have a choice, the enzyme preparation with lower K_m is preferable.

Because there are several species diffusing into and out of the gel, the normalization transformation must be done for all of them, leading to the system of second-order partial differential equations. Each species has its own Thiele modulus and again it is the smallest one that determines the overall outcome. The complicating factors are the reactions involving the buffer; these are very fast. In mathematical terms, it means that they are algebraic and the resulting partial differential equations are "stiff", requiring numerical solution (Caras et al., 1985a). As always with the differential equations, the final solution depends on the initial conditions and on the boundary conditions. The crucial ones define the conditions at the transducer/gel interface ($x = 0$ in Fig. 2.10). If none of the reacting species can cross this interface, the boundary is called a zero flux boundary. It is found in thermal, potentiometric, and optical sensors. However, in amperometric sensors, at least one of the species is consumed at this interface (typically oxygen), and a gradient of that species is established. The interface is then called the nonzero flux boundary. This difference in operating mechanism has a profound influence on the performance of such sensors, as we show in Chapter 7.

The boundary and initial conditions are always defined by the assumptions that have been made in the formulation of the model. These, in turn, depend
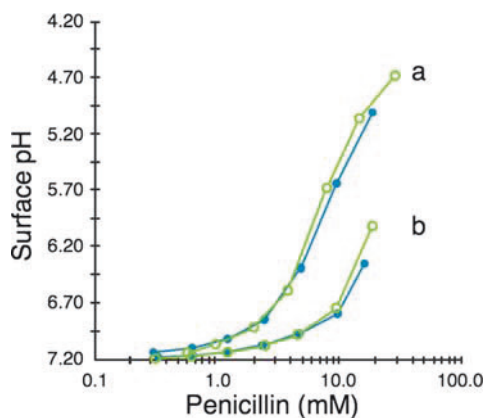
Fig. 2.10 Effect of buffer capacity on penicillin calibration curves: (a) 20 mM phosphate buffer and (b) 80 mM phosphate buffer (adapted from Caras et al., 1985a, p. 1925)

on approximations and compromises. Let us now review briefly the approximations that have been made, more or less historically, by various enzyme sensor investigators, and rank them in the approximate order of severity.

1. There is a linear diffusion gradient inside the enzyme layer.
2. There is no pH dependence of K_m.
3. There is no effect of mobile buffer capacity.
4. There is no effect of fixed (i.e., the gel itself) buffer capacity.
5. There is no partitioning of reactants and products between the gel and the sample.
6. There is no Donnan potential at the gel/sample boundary.
7. There is no depletion layer at the gel/sample boundary.

Approximations (1) and (2) have been made in the earliest models of development of enzymatic sensors in order to simplify the mathematics. They are both bad; the concentration profiles are nonlinear (Caras et al., 1985a; Eddowes, 1985) and the pH dependence of enzyme kinetics is an established fact.

Approximations (3) and (4) would be the most serious for enzymatic sensors in which the sensor output is related to the change of pH, because for such sensors the buffer capacity would have to be low and constant. However, for sensors that use some other reactants/products besides hydrogen ion, a large excess of buffer would mitigate the effects of these assumptions. To some extent, they can be also mitigated by the experimental design, as we show later.

The partitioning of electrically neutral species (assumption 5) and electrically charged species (assumption 6) between the gel and the sample affects the algebraic part of the model. It is a serious problem for both electrically neutral species (e.g., oxygen) and charged (ionic) species. However, it can, to some extent, be mitigated by the choice of the gel matrix.

Assumption (7) pertains exclusively to the enzymatic sensors with a nonzero flux boundary, at the gel/transducer interface (i.e., amperometric sensors). It can be eliminated by decreasing the size of the sensor.

In summary, assumptions (1) and (2) are unnecessary and have been avoided in more advanced models. Assumptions (3) and (4) are unavoidable and illustrate the fundamental weakness of most enzymatic sensors, particularly those depending on detection of pH changes. Assumptions (5) and (6) can be avoided to some extent by experimental design, but should be always accounted for in the model. Assumption (7) is easily avoidable. There is another assumption that has not been mentioned, the equality of concentration and activity. As discussed in Chapter 1, that cannot always be a justifiable assumption.

With these assumptions in mind, we now complete the outline of the solution of the diffusion–reaction problem as it applies to the most difficult case, the pH-based enzymatic sensors (potentiometric or optical). We assume only that there is no depletion layer at the gel/solution boundary (7), and that there is no fixed buffer capacity (4). The objective of this exercise is to find out the optimum thickness of the gel layer that is critically important for all zero-flux-boundary sensors, as follows from (2.26).

As a rule, hydrogen ion is involved not only in the pH-dependency of the reaction term (Thiele modulus) but also as the actively participating species involved in the acid–base equilibrium of all the substrates, reaction intermediates, products, and even the gel matrix. Furthermore, enzymatic reactions are always carried out in the presence of the mobile buffer. By "mobile" we mean a weak acid or a weak base that can move in and out of the reaction layer, as opposed to the fixed buffer represented by the gel (and by the protein) itself. Thus, we have to include the normalized diffusion–reaction equations for hydrogen ion and for the buffer.

$$\frac{\delta C_{H,T}}{\delta t} = D_H \frac{\delta^2 C_H}{\delta x^2} + D_{HA} \frac{\delta^2 C_{HA}}{\delta x^2} + \frac{v_{max} C_S}{\Re_{pH}(K_m + C_S)} \qquad (2.28)$$

Here $C_{H,T}$ is the total concentration of bound and unbound protons $\boldsymbol{H}$ within the enzyme layer and C_H is the concentration of protons. The term containing C_{HA} reflects the flux of the protonated buffer acid $\boldsymbol{A}$ in and out of the layer. The reaction term containing $\Re_{pH}$ is a characteristic property of the given enzyme and of the substrate and is discussed later. For simplicity, we consider here a simple monoprotic buffer for which

$$K_a = \frac{C_H C_A}{C_{HA}} \qquad (2.29)$$

Next, we have to define the boundary and the initial conditions. For the zero flux sensors (Fig. 2.9), the first space derivatives (i.e., fluxes) of all variables at the transducer/gel boundary (point $x = 0$) are zero:

$$\{C_S(0,t)\}'_x = \{C_H(0,t)\}'_x = \{C_{HA}(0,t)\}'_x = 0 \qquad (2.30)$$

On the other hand, in nonzero flux sensors the flux of at least one of the species (product or substrate) would be nonzero.

The concentrations of all species at the gel/sample boundary L are equal to the bulk values in the sample:

$$C_S(L,t) = C_{S,\text{bulk}}$$
$$C_{HA}(L,t) = C_{HA,\text{bulk}}$$
$$C_H(L,t) = C_{H,\text{bulk}}$$
$$C_A(L,t) = C_{A,\text{bulk}}$$

$$(2.31a\text{–}d)$$

The initial conditions are:

$$C_S(x,0) = 0 \; for \; x < L$$
$$C_{HA}(x,0) = C_{HA,\text{bulk}}$$
$$C_H(x,0) = C_{H,\text{bulk}}$$
$$C_A(x,0) = C_{A,\text{bulk}}$$

$$(2.32a\text{–}d)$$

This means that all species except the substrate are initially present inside the enzyme layer.

This treatment leads to a system of stiff, second-order partial differential equations that can be solved numerically to yield both transient and steady-state concentration profiles within the layer (Caras et al., 1985a). Because the concentration profile changes most rapidly near the $x = L$ boundary an ordinary finite-difference method does not yield a stable solution and is not applicable. Instead, it is necessary to transform the distance variable x into a dummy variable y using the relationship

$$y = L(1 - e^{-ax} + xe^{-a}) \tag{2.33}$$

This transformation allows for equal distribution in the y-space while concentrating the lines close to the $x = L$ boundary. Parameter a sets the spacing of the lines. This technique is called MOL1D (Method Of Lines in 1 Dimension) and is suitable for solving parabolic and hyperbolic initial boundary value problems in one dimension.

The actual solution for both transient and steady-state response of any zero-flux-boundary sensor can be obtained by solving (2.26) through (2.33) for the appropriate boundary and initial conditions. Fitting of the experimental calibration curves (Fig. 2.10) and of the time response curves (Fig. 2.11) to the calculated ones, validates the proposed model.

Once the theoretical curves have been fitted (Figs. 2.10 and 2.11), it is possible to plot the concentration profiles of all the species included in the model and to determine the optimum thickness of the enzyme layer (Fig. 2.12). Because the Thiele modulus is the controlling parameter in the diffusion–reaction equation, it is obvious from (2.22) that the optimum thickness will depend on the other constants and functions included in the Thiele modulus. For this reason, the optimum thickness will vary from one enzyme and one kinetic scheme to another.

Another important set of observations is related to the detection limit: dynamic range and sensitivity. For the expected values of the diffusion coefficient (in the gel) of approximately $10^{-6} \, \text{cm}^2 \, \text{s}^{-1}$ and substrate molecular weights about 300, the detection limit is approximately $10^{-4} \, \text{M}$. This is due to the fact that the product of the enzymatic reaction is being removed from the membrane by diffusion at approximately the same rate as it is being supplied. The dynamic range of the sensor
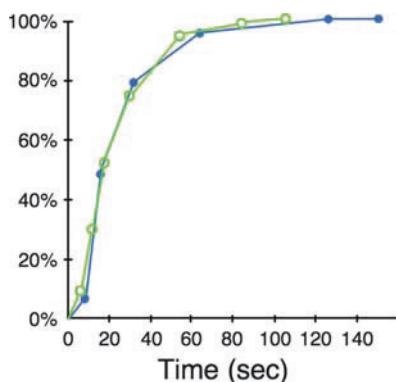
Fig. 2.11 Theoretical and experimental time response curve for penicillin (adapted from Caras et al., 1985a, p. 1925)
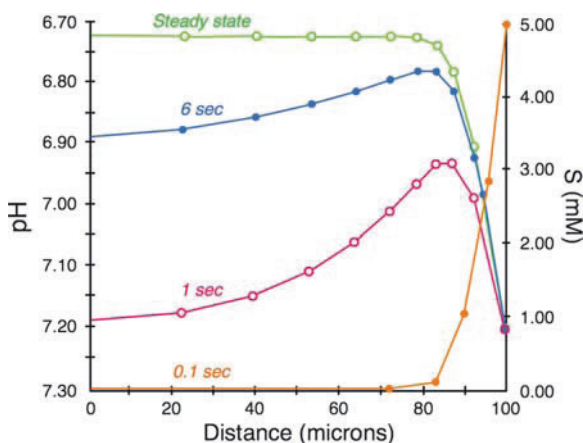
Fig. 2.12 Calculated concentration profiles for the substrate (penicillin) and the product (hydronium ion), expressed as pH (adapted from Caras et al., 1985a, p. 1918)

depends on the value of the K_m and on v_{max} (which depends on the enzyme loading). Generally speaking, higher loading should extend the dynamic range at the top of the concentration range. It is sometimes stated incorrectly that "the enzyme sensor has close to theoretical dependence" or a "Nernstian response," which means that a one-decade change of the bulk concentration of the substrate is expected to yield a one-decade change at the surface ($x = 0$) concentration. In the case of potentiometric enzyme sensors, it would yield a slope of approximately 60 mV/decade at 25°C. It is not intuitively obvious, but clearly evident from the comparison of the experimental and calculated response curves, that there is no general theoretical slope. Each enzymatic sensor has its own "theoretical curve," which depends on the mechanism and on the conditions under which it operates. We must remember that the decade/decade slope would occur only if a constant fraction of the product reached the $x = 0$ interface. The upper limit of the dynamic range depends on the value of the
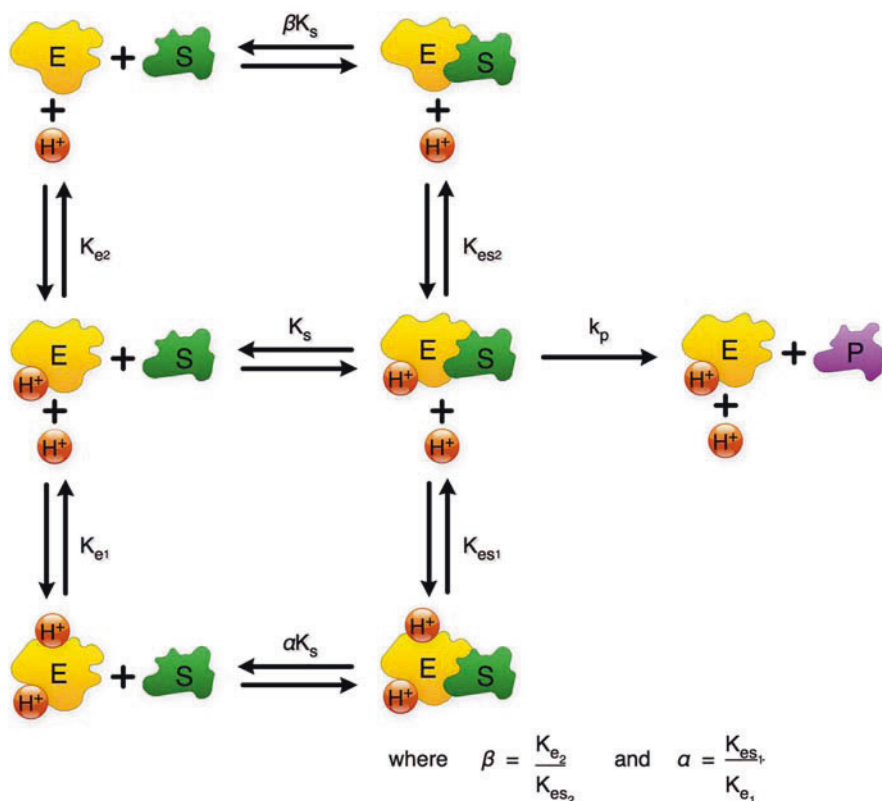
Fig. 2.13 One-enzyme substrate scheme, including protonation equilibria

Thiele modulus. It can be increased by the enzyme loading but, obviously, only up to a point. The normal dynamic range is approximately between 10^{-4} and 10^{-1} M.

We now return to the dimensionless pH-dependent reaction term $\Re_{\mathrm{pH}}$ in (2.28). Enzymes are proteins that are subject to multiple protonation equilibria. In that respect, they are polyelectrolytes. The scheme shown in Fig. 2.13 depicts the simplest situation, with only one product-forming pathway in which the product **P** is formed from the protonated enzyme/substrate complex $\mathbf{H^+ES}$.

Fractions of protonated and deprotonated enzyme are given by the dissociation equilibria, with appropriate dissociation constants. The substrate **S** shown in this scheme does not have acido-basic properties in the given pH range. Solving the equations outlined in Scheme 1 yields (for this reaction) the following relationship.

$$\Re_{\mathrm{pH}} = \left(1 + \frac{C_{\mathrm{H}}}{K_{\mathrm{ES}_1}} + \frac{K_{\mathrm{ES}_2}}{C_{\mathrm{H}}}\right) \tag{2.34}$$

This is the term that has to be inserted into (2.28) and normalized equations for all the other species involved in the reaction.

2.3.2 Zero-Flux-Boundary Sensors

Most enzyme sensor developers have used D-glucose/glucose oxidase as the first step in their studies. Glucose is an easily accessible, nontoxic substrate and D-glucose oxidase is an inexpensive and available enzyme. The reaction itself is amenable to thermal, electrochemical, and optical sensing. Blood glucose sensing is an important diagnostic problem (related to diabetes) that makes a good selling point in the perennial hunt for funding. Finally, some glucose oxidase-based sensors have been exceptionally commercially successful for specific diagnostic applications. Not surprisingly, there are thousands of glucose sensor papers in the literature. The current Internet entry "glucose sensor" netted 30,000 replies! Throughout this book, this reaction is used only for educational purposes, in order to highlight the important aspects of enzymatic sensing schemes. In no way does the inclusion of specific references imply endorsement of one or another approach.

Glucose oxidase belongs to a large and important family of enzymes that catalyze selective oxidation of various substrates. In nature, the obvious electron acceptor (oxidant) is oxygen which then becomes the second substrate in the kinetic scheme. In this form, the glucose oxidase has been used in many types of glucose sensors. Hydrogen peroxide is an intermediate in any reaction in which the oxygen is the ultimate electron acceptor. Because H_2O_2 is cytotoxic, another enzyme, catalase, always accompanies the natural oxidases. Its role is to remove the H_2O_2 as fast as it is formed. Nevertheless, a certain amount of hydrogen peroxide always escapes and causes damage to the parent oxidase, thus limiting its lifetime. This two-enzyme scheme is an example of an enzymatic cascade arrangement, in which the product of one enzymatic reaction (the intermediate) becomes the substrate for the next reaction. Quite often it is possible to base the sensing scheme on the interception of such an intermediate.

In spite of its importance and popularity, the fine details of the β-D-glucose oxidase mechanism are not completely known. The proposed model (Fig. 2.14) includes both the catalase cascade and the protonation equilibria (Caras et al., 1985b). The pH-dependent reaction term corresponding to this model is quite complex.

$$\Re_{pH} = \left[\frac{1 + \frac{C_H}{K_5} + \frac{K_{5'}}{C_H}}{k_{cat}C_{E_{total}}} + \frac{\frac{C_H}{K_3} + 1}{k_3 C_S C_{E_{total}}} + \frac{\frac{K_4}{C_H} + 1}{k_4 C_{O_2} C_{E_{total}}} \right]^{-1} \qquad (2.35)$$

The verification of the model is again performed by fitting the experimental calibration (Fig. 2.15) and time response (Fig. 2.16) curves.

The fits in this case are not as good as the ones obtained for the penicillin case (Figs. 2.10 and 2.11). This is due to the fact that the glucose oxidation mechanism is not yet completely understood and the kinetic equations are only approximate. Nevertheless, it is again possible to plot the profiles of the most important species in the gel layer and from this fit to estimate the optimum thickness of the gel layer (Fig. 2.17). For the glucose sensor, the optimum thickness appears to be 150 μm,
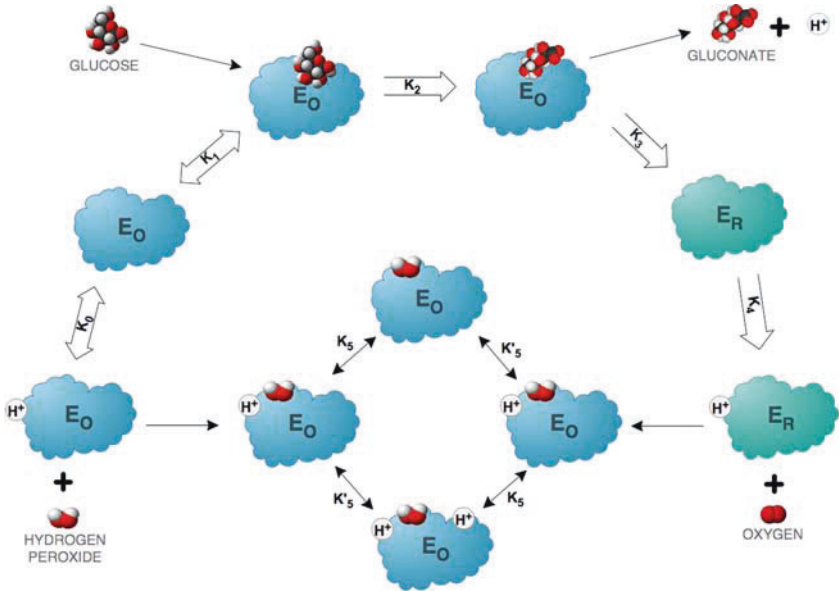
Fig. 2.14 Glucose oxidation scheme
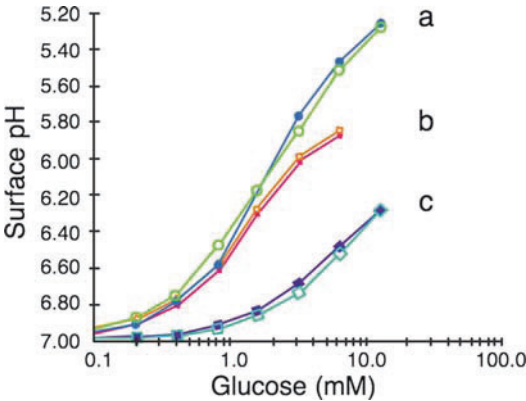
Fig. 2.15 Theoretical (open points) and experimental (full points) calibration curve for glucose sensor for (a) 100% oxygen in 0.2 mM buffer, (b) 25% oxygen and 0.2 mM buffer, and (c) 100% oxygen and 1 mM buffer (adapted from Caras et al., 1985b, p. 1921)

under the loading conditions given earlier in this section. The most important result of this procedure is to estimate the effect of the buffer capacity and of oxygen concentration on the pH at the hydrogel/transducer boundary. This result clearly indicates that both the buffer capacity and oxygen are serious interferences and practically negate the high selectivity of the enzyme itself. This is the most serious reason why zero-flux-boundary sensors (i.e., potentiometric or optical) have failed in all but the most well-defined laboratory conditions.
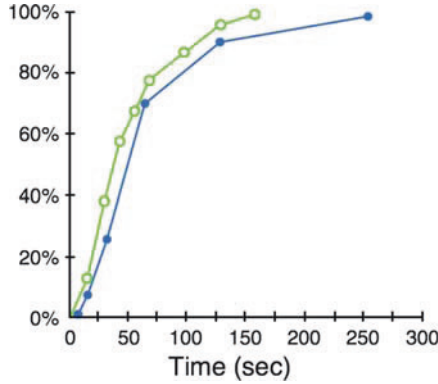
Fig. 2.16 Theoretical (full points) and experimental (open points) time response curve for glucose sensors to step change in concentration (from 0 to 1 mM) (adapted from Caras et al., 1985b, p. 1922)
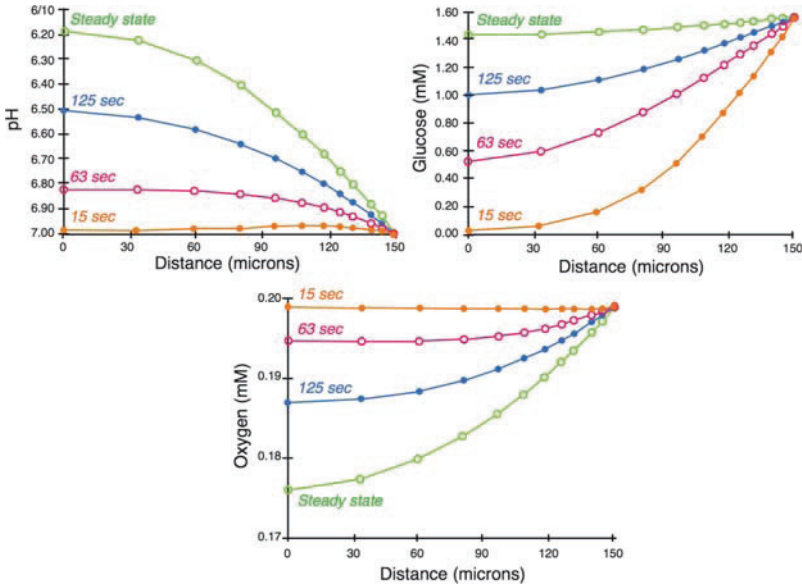
Fig. 2.17 Evolution of concentration profiles, calculated from Figs. 2.15 and 2.16 (adapted from Caras et al., 1985b, p. 1922)

2.4 Mass Transport Selectivity

This form of selectivity is based on the concept of selectively blocking the access of all interfering species to the active region of the transducer. It is a form of filtration. The blockage can be achieved by size discrimination. For instance, a dialysis

membrane placed in front of a mixture containing the transducer species and other species can selectively filter out all species above a certain cutoff size. Smaller species (and presumably only the transducer species) can be allowed in.

This may work well if the process involves only electrically neutral species. However, when ions are discriminated on the basis of size, the partitioning process is affected by the Donnan potential. This potential, which we discuss more fully in Chapter 6, develops at the membrane/electrolyte interface. Another possibility is to discriminate on the basis of charge, as shown in Fig. 7.10 (see Chapter 7). Again, a porous barrier membrane is used, although here it would contain fixed, electrically charged moieties. When placed in front of the transducer, it rejects the like-charged species by electrostatic repulsion. In other words, it is a form of ion exchange membrane.

2.5 Design of Selective Layers

The selection, preparation, and properties of a selective layer depend largely on the type of transducer at which they will be used, as well as on the application. Those aspects are discussed in the context of the individual transduction principles. Only certain common features and procedures are included in this section.

With only a few exceptions, such as enzyme-containing layers and some ion-selective electrode membranes, the selective layers are on the order of a few micrometers thick. Therefore, common thin-film preparation techniques can be used, particularly if the uniformity of preparation is important. For layers prepared from solvents, spin-coating is the preferred technique, because it offers good control of both the thickness and uniformity. Dip-coating and drop-casting are often used for preparation of individual sensors. The rate of evaporation of the solvent influences the porosity and density of the film. It is preferable to evaporate the solvent slowly either by choosing a higher boiling point solvent, or evaporating it in an enclosed compartment against finite vapor pressure of the solvent. This is particularly important to prevent formation of the "skin," and to achieve better adhesion to the substrate.

Vacuum deposition techniques, such as sputtering, electron beam evaporation, and plasma deposition are common. Photopolymerization and laser-assisted depositions are used for preparation of specialized layers, particularly in the fabrication of sensing arrays. Most commercial instruments have thickness monitors (Chapter 4) that allow precise control of the deposition process.

2.5.1 Preparation of the Substrate

Preparation of the substrate at which the layer is deposited is critically important. Some deposition techniques operate at an elevated temperature of the substrate

and/or a thermal step is included somewhere in the fabrication sequence, for example, wirebonding. The difference of the thermal expansion coefficients of the selective layer and of the substrate must always be considered. If the mismatch is too severe, delamination occurs at the substrate/layer interface and leads to device failure.

To some extent, it can be mitigated by the chemical preparation of the substrate. The first step is the removal of *dirt* ("matter in the wrong place") which may have inadvertently contaminated the surface during one of the preceding steps. Oxygen plasma cleaning (*ashing*) generally removes organic residues. Aggressive liquid cleaners such as the "piranha solution" (H_2O_2/H_2SO_4), can also be used, but with appropriate safety precautions. Rinsing with deionized, organics-free water is usually the final step. Use of acetone is generally not recommended because it is rarely available in sufficient purity and often leaves a thin organic residue on the substrate.

The second step is adhesion. The use of adhesion promoters is common. For oxide surfaces, silanization is by far the most popular. It can also be used for introduction of specific binding sites to the surface. It is based on the following reaction of surface hydroxyl groups with one or more reactive groups of the silane derivative, shown at the bottom of Table 2.3. Thus, a multipoint attachment, as well as introduction of the desired functional group **X** to the substrate surface, is achieved. There are many different "home kitchen" recipes that are used for the silanization. An important thing to remember is that the chloro-groups on the silane are much more reactive than the alkoxy-groups. Therefore, the reaction conditions must be adjusted accordingly. The objective of this procedure is to achieve surface activation, meaning that a monolayer of the functional groups should be the ideal result. If the silanization reaction is allowed to proceed for too long, particularly for chlorosilanes, a multilayer siloxane layer is formed that may interfere with the sensing function. Also, for the silane to react it must have an active surface hydroxyl group available. If the immediately preceding fabrication step involved temperatures above $\sim$200 °C there may not be a sufficient number of –OH groups at the surface, resulting in poor adhesion. Therefore, a brief exposure of the oxide surface to water, followed by air-drying to remove the excess, is recommended.

Another common adhesion promotion scheme is used when thin layers of noble metals such as gold or platinum are used in the sensor. Because noble metals do not have a high affinity for oxide surfaces, their adhesion is poor and they often delaminate, particularly when exposed to solutions. Intermediate "glue metals" are used to overcome this problem. The most common mistake made is the use of chromium for this purpose. It has long been known – and just as long ignored – that chromium migrates along the grain boundaries of the deposited noble metal (Holloway, 1979). When it reaches the surface, it reacts with oxygen, covering the entire surface of the metal with Cr_2O_3. This process is surprisingly fast, and it is assisted by elevated temperature. It takes only a few hours for Cr to travel through several hundred nm of Au. The problem is less severe but also present in Pt (Josowicz et al., 1988). Chemical cleaning of the chromium-contaminated surface offers only temporary relief because the Cr migration continues. Therefore, use of Cr should be avoided. An acceptable alternative is to use a thin ($\sim$20 nm) layer of Ti. It is a more reactive

Table 2.3 Reactive functional groups (adapted from Pace, 1981)

Reactive group I (e.g. on surface)	Intermediate	Reactive group II (e.g. on terminal reagent)	Coupling linkage "type"
-R-NH$_2$	R'-N-C-N-R" (Carboniimide)	HO-C- (O=C)	-HN-C- (O=C) Amide
–NH$_2$ (phenyl)	-N$_2$Cl (phenyl, ⊕ ⊖) Diazonium	HO– (phenyl)	HO–(phenyl)–NH=N–(phenyl)–HO Diazo
–NH–NH$_2$ (phenyl)		-C- (O=C)	–(phenyl)–NH-N=C- Hydrazide
-SH		HS-	–S-S– Disulfide
-SH	R'-N=C-N-R"	HO-C- (O=C)	–S-CO– Thioamide
Maleic anhydride		–NH$_2$–	(structure) C—C(=O)OH / C—C(=O)NH–
–COOH	-C-CH$_2$ (O=C)	–NH$_2$–	–CONH– Amide
–CHO	Schiff base	– H$_2$N–	–C=N–
C–OH / C–OH	–CNH–	– H$_2$N–	C–OCO-NH– / C–OH
(triazine Cl, N, N, N, R)	Cyanuric chloride	–H$_2$N–	(triazine NH–, N, N, N, R)
OH / OH / OH	Cl,Cl,Cl— Si-(CH$_2$)$_3$-X		O,O,O— Si-(CH$_2$)$_3$-X
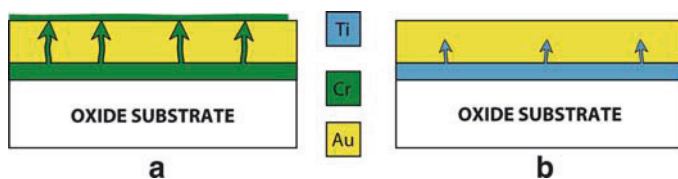
Fig. 2.18 Adhesion promotion on noble metal substrates

metal than Cr and forms a strong bond with the oxide. The next Ti layer then alloys with the noble metal. Most important, the migration of Ti along the grain boundaries stops as soon as Ti is even partially oxidized. An interface structure that works for both Au and Pt is shown in Fig. 2.18.

Surface characterization techniques, such as photoelectron spectroscopy, can be used to verify the quality of the surface of such layers.

2.5.2 Immobilization of Specific Binding Sites

This section pertains particularly to proteins and other biomolecules that are used as specific binding sites in biosensors. For that purpose, these molecules have to be retained in the selective layer or at the selective surface of the sensors. Because they are usually large, one possibility is to use entrapment in a matrix. Both organic and inorganic matrices have been used for this purpose (Lev et al., 1995). However, this is not a preferred approach due to the poor control of the porosity of the matrix.

Therefore, some form of covalent attachment is preferred. In this case, it is important not to destroy the functionality of the biomolecule by affecting the immobilization at or close to its active center. This is generally achieved by using a two-step immobilization approach. In this scheme, either the surface or the matrix is first activated by introduction of some reactive functional group, and is then reacted with the linking functional group on the surface of the biomolecule. To that end, aliphatic amino groups (e.g., on lysine or arginine), thiol, or carboxylic groups have been used.

A simple, one-step immobilization technique is bulk cross-linking of the functional protein with bovine serum albumin (BSA) using glutaraldehyde as the cross-linking agent (Fig. 2.19). It is popular because of its simplicity, but it usually leads to reduction of the biological activity of the biomolecule. A cleaner and preferable approach is the two-step carbodiimide route. A partial summary of the various immobilization options was shown in Table 2.3.
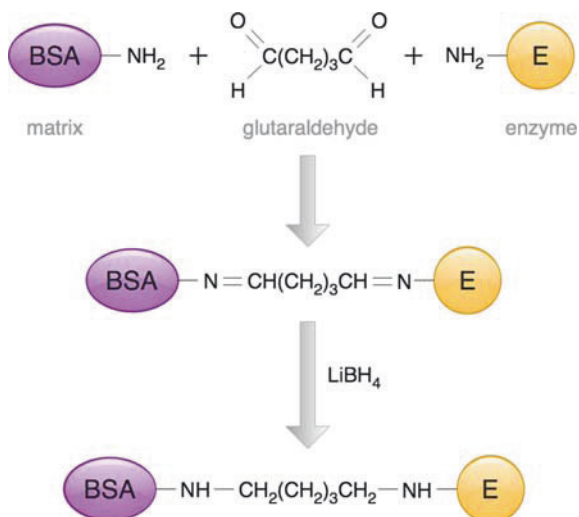
Fig. 2.19 Glutaraldehyde cross-linking of enzyme E with bovine serum albumin (BSA), followed by Li borohydride stabilization

Food for Thought #2

DNA Selectivity

The formation of a duplex between fully complementary ssDNA is one of the most selective and strongest interactions between two molecules. The matching of A–T and G–C is driven by the highly specific stereochemistry of hydrogen bonding. As a result, the binding constant resulting from such multiple interactions is exceptionally high, typically greater than $10^{10}\,M^{-1}$. It has caught the attention of sensor people and tens of papers based on DNA sensing have been published, utilizing gravimetric (QCM), optical, and electrochemical transduction principles.

1. Is the selectivity of the individual A–T/G–C pair matching reflected in the overall selectivity of such sensors? Why yes or no?
2. Probe and target DNA sequences in which 10% of the bases are mismatched are effectively a "different species." In this scenario, would the target still bind to the probe DNA?
3. Under what conditions could the high hybridization selectivity be fully utilized for direct (reversible) sensing?
4. How does the presence of 5%, 10%, and 20% mismatches affect the performance of, for example, QCM sensors?

Adsorption and Absorption

One weakness of chemical sensors based on adsorption of analyte at their surface is their poor selectivity, in other words, their vulnerability to interference from other species. Contrast this situation with the sensors having binding sites buried in the bulk of the selective layer. Such sites are accessible only by absorption.

The "surface" of a porous solid is a relative notion, dependent on the size of the species. There are porous selective layer matrices so "open" that small species can easily diffuse into their bulk without any specific interaction, yet they filter out large analytes.

5. How and where would you utilize this notion in the optimization of performance of a selective layer and what would be the trade-offs?

Gas Immunosensors

Antibodies against small molecules (haptens) can be readily prepared. It has been suggested that Ab against, for example, pesticide gas molecules can be immobilized on mass sensors and used for sensing of such gaseous compounds.

6. What will be the major interferant in such scheme?
7. Would this problem be encountered for any biosensor?
8. Assume that the hydrophobic bond is a major component of the binding interaction of the gas with the antibody binding site. What is the consequence for that type of chemical bond if performing this reaction is carried out in gas phase?

Linear Solubility Energy Relationship

Linear Energy Solubility Relationships (LSERs) are useful in the design of selective layers for mass and optical gas sensors.

9. Imagine that a polymer has been fully characterized and its free energy contribution coefficients for individual vapors A and B have been evaluated according to the LSER equation. If the polymer is fully saturated with vapor A, do you expect the coefficients for vapor B to be the same as for the polymer in the absence of vapor A? Would the evaluation of the retention times for A and B on a GC column packed with the same polymer offer any help in such a case?
10. Contrast the difference in interaction of the mixture of vapors with the solid phase used in a gas chromatographic experiment and in a direct sensing application.

Molecular Imprinting

Molecular imprinting requires multiple short-range weak interactions that act over short distances. It means that the molecular fit of the template/analyte pair must be very tight.

11. What is the effect of this requirement on removal of the template molecule from the imprint and the access of the analyte to the imprint. If it is a problem, suggest how it could be circumvented. How does Nature deal with the tight fit problem?
12. Some MIPs are created by solution polymerization. This means that the template molecule is always solvated, to some extent. Consider the role that such solvation would play in creating the imprint that is expected to mirror only the nonsolvated template molecule.
13. Are the solvation issues similar or different in the aptamer footprint forming and applications? Explain why yes or no.
14. How would you design a correct control experiment to verify that the imprinting really works?

Symbols

$a\Sigma\alpha_2^H$	Term defining hydrogen bonding at an acidic site
$b\Sigma\beta_2^H$	Term defining hydrogen bonding at a basic site
C	Concentration
D	Dielectric constant
ΔG	Change in free energy
ΔH	Change in enthalpy
I	Ionization energy
K	Binding constant/equilibrium constant/partitioning coefficient
K_m	Michaelis–Menten constant
k	Rate constants
L	Thickness of a layer
$\Re_{pH}$	Dimensionless pH-dependent activity of enzyme
r	Distance
rR_2	Polarizability term
ΔS	Change in entropy
$s\pi_2$	Polarity term
t	Time
v	Reaction velocity
v_{max}	Maximum reaction velocity
x	Distance
z	Charge
α	Polarizability
δ	Fraction of transferred charge

μ	Dipole moment
Θ	Angle
ϕ	Thiele modulus

References

Abraham, M.H. (1993) *Chem. Soc. Rev*. 22, 73.

Abraham, M.H., Andonian-Haftvan, J., Chau My Du, Diart, V., Whiting, G.S., Grate, J.W., and McGill, R.A. (1995) *J. Chem. Soc. Perkin Trans*. 2, 369.

Absolom, D.R. and van Oss, C.J. (1986) *CCC Crit. Rev. Immunol*. 6, 1.

Caras, S.D. and Janata, J. (1985) *Anal. Chem*. 57, l924.

Caras, S.D., Janata, J., Saupe, D., and Schmitt, K. (1985a) *Anal. Chem*. 57, 1917.

Caras, S.D., Petelenz, D., and Janata, J. (1985b) *Anal. Chem*. 57, l920.

Diaz-Garcia, M.E. and Badia, R. (2004) Molecularly imprinted polymers for optical sensing devices. In: O.S. Wolfbeis (Ed.), *Optical Sensors*. Springer.

Eddowes, M.J. (1985) *Sens. Actuat*. 7, 97.

Ekedahl, L.-G., Eriksson, M., and Lundstrom, I. (1998). *Acc. Chem. Res*. 31, 249.

Grate, J.W., Patrash, S.J., Abraham, M.H., and Chau My Du (1996) *Anal. Chem*. 68, 913.

Haupt, K. (2004) Molecularly imprinted polymers as recognition elements in sensors. In: O.S. Wolfbeis (Ed.) *Ultrathin Electrochemical Chemo-and Biosensors*. Springer.

Hierlemann, A., Zellers, E.T., and Ricco, A.J. (2001) *Anal. Chem*. 73, 3458.

Holloway, P.H. (1979) *Gold Bull*. 12, 99.

Huaiqiu Shi, Tsai W.B., Garrison, M.D., Ferrari, S., and Ratner, B.D. (1999) *Nature* 398, 593–597.

Janata, J. and Josowicz, M. (1998) *Acc. Chem. Res*. 31, 241–248.

Janeway, C., Travers, P., Walport, M., and Shlomchik, M. (2004) *Immunobiology: The Immune System in Health and Disease*, 6th ed. Taylor & Francis.

Josowicz, M., Janata, J., and Levy, M. (1988) *J. Electrochem. Soc*. 135, 112.

Lev, O., Tsionski, M., Rabinovich, L., Glezer, V., Sampath, S., Pankratov, I., and Gun, J. (1995) *Anal. Chem*. 67, 22A–30A.

Pace, S.D. (1981) *Sens. Actuat*. 1, 475.

Reichardt, C. (1988) *Solvents and Solvent Effects in Organic Chemistry*, 2nd ed. VCH.

Tombelli, S., Minunni, M., and Mascini, M. (2005) *Biosens. Bioelectron*. 20, 2424–2434.

Topart, P. and Josowicz, M. (1992) *J. Phys. Chem*. 96, 7824–7830; 8662–8666.

Tuerk, C. and Gold, L. (1990) *Science* 249, 505–510.

van Holde, K.E. (1985) *Physical Biochemistry*, 2nd ed. Prentice-Hall.

Chapter 3
Thermal Sensors

3.1 Introduction

The first law of thermodynamics tells us that any process in which the internal energy of the system changes is accompanied by absorption or by evolution of heat. The class of chemical sensors discussed in this chapter uses the heat generated by a steady-state chemical reaction as the source of analytical information.

3.2 General

Thermal sensors utilize kinetic selectivity, therefore, some form of catalysis is always involved. The important point to realize is that thermal sensors are in situ microcalorimeters, which means that batch calorimetry can provide important information for thermal chemical sensors.

There are two properties of heat that are quite unique from any other physical parameter. First, heat is totally nonspecific. Second, it cannot be contained; that is, it spontaneously flows from the warmer (T_1) part of the system to the colder (T_2). From the sensing point of view, this defines the optimal design of a thermal sensor. The general strategy is to place the chemically selective layer on top of a thermal probe and measure the heat evolved in the specific chemical reaction taking place in that layer, either as the change in temperature of the sensing element or as the heat flux through the sensing element. Because the heat is evolved continuously, thermal sensors are, by definition, nonequilibrium sensors.

At constant pressure, the generated heat is equal to the enthalpy change ΔH, which is related to change of the temperature of the closed system. This is called the adiabatic condition.[1]

$$\Delta T = \frac{\Delta H}{C_{\mathrm{p}}} \tag{3.1}$$

---

[1] See (A.7) and (A.9) for more thermodynamic background.

where C_p is the heat capacity at constant pressure. If C_p is constant over a given temperature range, the enthalpy change can be calculated from this measurement.

Another useful design parameter is the thermal capacity (C_s), which is heat capacity divided by the mass m.

$$C_s = \frac{C_p}{m} \qquad (3.2)$$

The thermal capacity is the amount of heat needed to raise the temperature by one degree. Its units are $J\,kg^{-1}K^{-1}$. It is obvious that in the design of thermal sensors we want to use materials that have a low thermal capacity.

3.3 Push–Pull Operation

Another way to maximize the overall performance of thermal sensors is the push–pull principle that is used in isothermal calorimetry. In that case, the heats flowing into and out of the reactor are exactly balanced, so that the temperature inside the reactor remains constant.

$$\frac{d}{dt}(\pm\Delta H_{rxn} + \Delta H_{el} + \Delta H_{loss})_T = 0 \qquad (3.3)$$

Equation (3.3) shows the individual heats that contribute to overall heat gain and loss. The reaction heat, ΔH_{rxn}, is positive for endothermic reactions and negative for exothermic reactions. The heat loss to the environment is shown as ΔH_{loss}. The heat supplied to the reactor, usually in the form of electric heating, is ΔH_{el} but other forms of heating, for example, radiative, can be used instead of the electric heater.

The baseline condition applies when no chemical reaction takes place.

$$\frac{d\Delta H_{rxn}}{dt} = 0 \qquad (3.4)$$

When the heat lost to the environment is exactly balanced by the heat supplied by the heater, (3.5) is obtained.

$$\left(\frac{d\Delta H_{el}}{dt}\right)_T = -\left(\frac{d\Delta H_{loss}}{dt}\right)_T \qquad (3.5)$$

For a Joule-resistive electrical heating element

$$\left(\frac{d\Delta H_{el}}{dt}\right)_T = IV = I^2R \qquad (3.6)$$

Here, V is voltage, I is current, and R is resistance. When a chemical reaction takes place and the temperature in the environment is constant, the thermal output of the chemical reaction is directly related to the electric heater output.

$$\pm \left(\frac{\mathrm{d}\Delta H_{\mathrm{rxn}}}{\mathrm{d}t} \right)_T = \mp \left(\frac{\mathrm{d}\Delta H_{\mathrm{el}}}{\mathrm{d}t} \right)_T \tag{3.7}$$

Thus, for an exothermic ($\Delta H_{\mathrm{rxn}} < 0$) chemical reaction, the heater output is reduced, whereas for an endothermic ($\Delta H_{\mathrm{rxn}} > 0$) chemical reaction, the heater output is increased, in order to maintain the thermal steady-state. Obviously, for this scheme to work, the temperature of the environment has to be constant and lower than the temperature of the reactor, so that $\Delta H_{\mathrm{loss}} < 0$. This condition may be difficult to achieve for thermal biosensors that operate at or near room temperature. In that case, it is possible to use a constant output electrical cooler (e.g., a Peltier element). The electrical heater then makes up for the constant heat lost due to the cooling. The performance of the sensor improves as the temperature differential between the sensor and its environment grows larger. For operation with external cooling, a constant cooling term is added to (3.3). The push–pull scheme is equally suited for thermal sensors operating in gas and in the liquid phase, where the thermal capacity of the surrounding medium is higher. Therefore, both biocatalysis and gas phase catalysis can be used in the design of push–pull thermal sensors.

3.4 Thermometers

The essential part of the thermal chemical sensor is the device that allows fast, sensitive, and precise detection of the temperature: a thermometer. There are many thermometers available; here we mention only three types, in the descending order of their sensitivity. They are thermocouples, platinum resistors, and thermistors.

3.4.1 Platinum or Platinum–Iridium Alloy Resistors

The Pt resistor is a thin wire whose resistivity ρ changes with temperature, according to the equation

$$\rho_T = \rho_0(1 + AT + BT^2) \tag{3.8}$$

The coefficients for Pt are $A = 4 \times 10^{-3}$, $B = 5.8 \times 10^{-7}$, and $\rho_0 = 1 \times 10^{-5}$ Ω cm. With these parameters, the sensitivity, expressed as the temperature coefficient, is $\sim 0.4\% \,^{\circ}\mathrm{C}^{-1}$ over a wide range of temperatures. Resistivities of other metals, as well as their temperature coefficients, are tabulated in standard reference tables (e.g., the CRC Handbook of Chemistry and Physics, 2006). Because the geometry of the resistor does not change with temperature, (3.8) is often written in terms of change of resistance R. Because of their chemical inertness and high temperature coefficient, platinum resistors are most common. They are the key part of the most successful thermal sensors, pellistors, which are discussed in Section 3.6.2.

3.4.2 Thermocouples

Thermocouples operate on the principle of thermoelectricity (Seebeck effect). When the junction of two dissimilar electronic conductors – typically two metals, two semimetals, or semiconductors – is held at temperature T_1 (the so-called "cold junction"), a measurable voltage develops between it and the "hot" (T_2) junction. This thermoelectric voltage (V) of the two conductors can be measured according to (3.9).

$$V = (Q_A - Q_B)(T_1 - T_2) \qquad (3.9)$$

The Seebeck coefficients Q_A and Q_B are material constants of conductors **A** and **B**, respectively. They depend primarily on two parameters: their work function (see Appendix C) and their thermal conductivity. There are many combinations of electronic conductors producing V of few mV $°C^{-1}$. It is interesting to note that direct modulation of one or both Seebeck coefficients by chemical interaction with an electron acceptor or electron donor gas is possible. It has been demonstrated as a sensing principle for detection of gaseous NO_2 with an In_2O_3/Au thermocouple junction (Liess and Steffes, 2000).

3.4.3 Thermistors

These are the probes of choice in thermal chemical sensors because they are stable ($\pm 0.05\,°C$/year), inexpensive, sensitive, chemically inert, and small (Herwaarden and Meijer, 1994). Moreover, they can be directly heated and as such can be used for the push–pull operation. They are made of high-temperature-sintered oxides (e.g., BaO/SrO) covered with glass coating (Fig. 3.1).

Thermistors cover a temperature range from $-80\,°C$ to $+350\,°C$, and have either a negative (NTC) or positive (PTC) temperature coefficient. Typical resistance ranges from around 100Ω up to the megaohm (MΩ) region. Their response is related
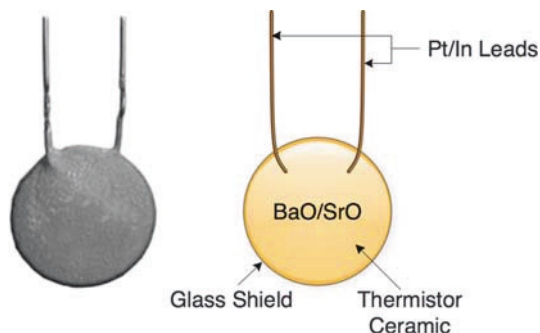
Fig. 3.1 Directly heated, shielded thermistor

to the bandgap energy, E_g, of the oxide semiconductor:

$$R_T = R_0 e^{(\pm E_g/2kT)} \tag{3.10}$$

Equation (3.10) can be differentiated to yield (3.11).

$$\frac{dR_T}{dT} = \pm \frac{R_0 E_g}{2kT^2} e^{(\pm E_g/2kT)} \tag{3.11}$$

The relative change of resistance is its sensitivity characterized as the thermistor coefficient α. It is obtained by dividing (3.10) by (3.11).

$$\frac{dR_T}{R_T dT} = \alpha = \pm \frac{E_g}{2kT^2} \tag{3.12}$$

Thus, the sensitivity of a thermistor is a quadratic function of operating temperature. The optimum sensitivity is typically $\pm 4\%$ $°C^{-1}$. Because of their nonlinear response, thermistors are sometimes linearized by placing a resistor of similar nominal value in parallel with the thermistor, with a resulting loss of sensitivity. This is not generally necessary for thermochemical sensors, particularly for the push–pull applications, because the temperature range involved is small. For direct temperature measurement, the detection limit of $10^{-4} °C$ can be achieved with a conventional Wheatstone bridge.

3.5 Equivalent Circuits

Modeling and optimization of chemical sensors can be assisted by creating equivalent electrical circuits in which an ordinary electrical element, such as a resistor, capacitor, diode, and so on, can represent an equivalent nonelectrical physical parameter. The analysis of the electrical circuit then greatly facilitates understanding of the complex behavior of the physical system that it represents. This is a particularly valuable approach in the analysis and interpretation of mass and electrochemical sensors, as shown in subsequent chapters. The basic rules of equivalent circuit analysis are summarized in Appendix D. Table 3.1 shows the equivalency of electrical and thermal parameters that can be used in such equivalent circuit modeling of chemical thermal sensors.

3.6 Examples of Thermal Chemical Sensors

A schematic diagram of a sensor based on a thermistor is shown in Fig. 3.2. Because of their sensitivity, thermistors are preferred over thermocouples; however, any type of thermometer can be used in such design. There are some conflicting requirements

Table 3.1 Electrical and thermal equivalency (data from Herwaarden and Meijer, 1994, p. 331)

Thermal parameter	Electrical parameter
Temperature T (K)	Voltage V (Volt)
Heat flow, power P (W)	Current I (Ampere)
Heat ΔH (J = Wt)	Charge q (Coulomb)
Resistance R	Resistance R (Ohm)
Conductance G (W/K)	Conductance G (Siemens)
Heat capacity C_p (J/K)	Capacitance C (Farad/Volt)
Thermal conductivity κ (W/K-m)	Electrical conductivity σ (S/m)
Thermal mass C_s (J/kg-K)	Permittivity ε (F/m)
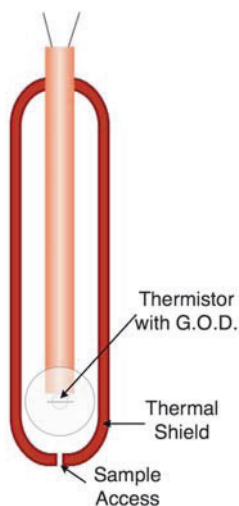
Fig. 3.2 General scheme of thermistor-based enzyme biosensor

built into the operation of this device: in order for the sensor to interact with the chemical species in the sample, the access of the sample to the selective layer must be allowed; that is, it must be a thermodynamically open system. On the other hand, in order to obtain a maximum response it should be as adiabatic (i.e., as thermally isolated) as possible. These conflicting requirements present some challenge in the optimum design configuration.

3.6.1 Glucose Thermistor

In order to provide a steady-state signal, heat must be continuously evolved. An ideal reaction for this is the biocatalytic enzymatic reaction, which combines high substrate specificity with a high amplification factor. Thus, an enzyme-containing layer is deposited over the thermal probe, and the substrate is allowed to diffuse in.

As it diffuses in, it reacts according to a general Michaelis–Menten equation (2.21), and the molar heat equal to the enthalpy of that reaction is evolved. For an enzymatic reaction

$$E + S \Leftrightarrow ES \Rightarrow P \quad -\Delta H_{rxn} \tag{3.13}$$
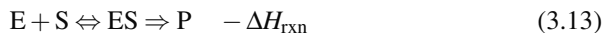
This is a typical diffusion–reaction scheme, as discussed in Chapter 2 (2.13). The major difference is the unconfined heat flow, which makes the solution of the response equation, even by digital simulation, difficult.

A further complicating factor is the fact that every reaction (including the protonation equilibria) contributes to the overall thermal effect. The corollary of the anisotropic heat flow is that the overall temperature change in the thermal sensing element is relatively small (typically 0.1–1 m°C) because only a fraction of the total heat contributes to the useful temperature change. Inevitably, response curves for sensors of this type are always empirical. The glucose-sensitive thermistor, using immobilized glucose oxidase (GOD; Tran-Minh and Vallin, 1978), illustrates the concept and the problems. The reactions in which heat is evolved are the same as shown in Chapter 2.

$$\beta\text{-D-glucose} + H_2O + O_2 \xrightarrow{GOD} H_2O_2 + \text{D-gluconic acid} \quad \Delta H_1 \tag{3.14a}$$

$$H_2O_2 \xrightarrow{catalase} \frac{1}{2}O_2 + H_2 \quad \Delta H_2 \tag{3.14b}$$

The total enthalpy change is the sum of the two partial enthalpies ΔH_1 and ΔH_2, approximately $-80\,\text{kJ mol}^{-1}$. The enzymes are immobilized on the tip of the thermistor, which is then partially enclosed in a glass jacket in order to reduce the heat loss to the surrounding solution (Fig. 3.2). A second thermistor with similarly immobilized bovine serum albumin is used as a reference device in the opposing arm of the Wheatstone bridge. The response of this arrangement is shown in Fig. 3.3. The detection limit is m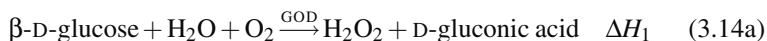odest: $\sim$3 mM. The problem clearly lies in the high heat capacity and high heat conductivity of water. Other enzyme thermistors with similar performance characteristics and problems have been made using urease and other enzymes.

3.6.2 Catalytic Gas Sensors

The mass/heat flow conflict is considerably simpler in gas sensors, due to the much lower heat capacity of gases compared to water. These sensors are known under the generic name "pellistors" (Dabill et al., 1987). Here again, the heat is liberated at the surface of the sensor by the catalytic reaction; the resulting temperature change inside the device is measured. With the safety of mining operations in mind, catalytic gas sensors have been designed specifically for the detection of minute concentrations of flammable gases in ambient air. These subthreshold concentrations are beneath the Lower Explosion Limit (LEL) for the applicable gas. The chemistry
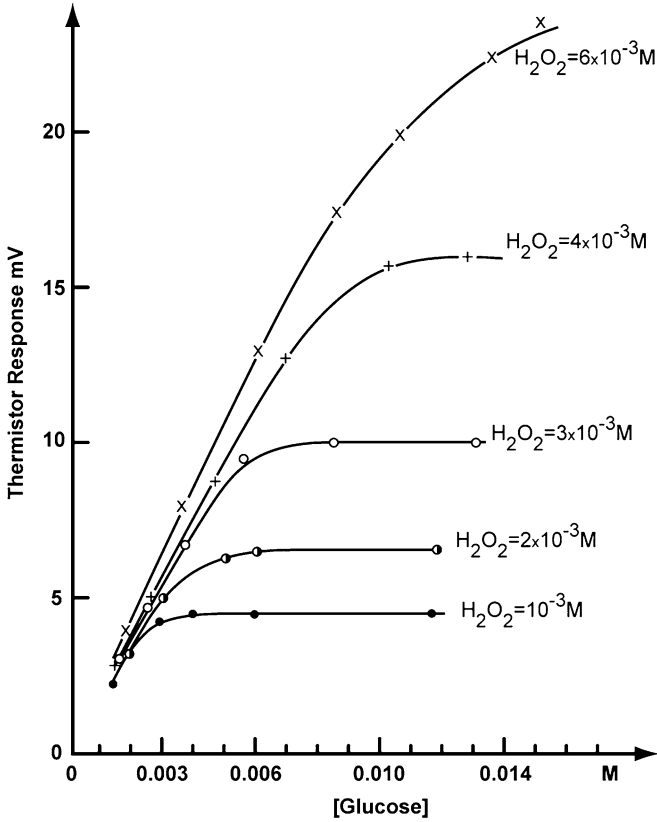
Fig. 3.3 Response of glucose oxidase thermistor to glucose in the presence of different concentrations of hydrogen peroxide (adapted from Tran-Minh and Vallin, 1978, p. 1874)
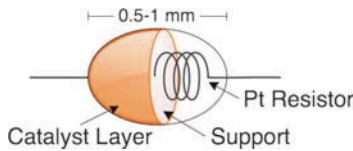
Fig. 3.4 Schematic of a pellistor

that is involved is very similar to that of high-temperature conductometric oxide sensors (Chapter 8). They are one of the simplest sensors available (Fig. 3.4).

The platinum coil is embedded in a pellet of sintered ceramic material, such as ThO_2/Al_2O_3, which is then coated with porous catalytic metal, for example, palladium or platinum. Other metals have been also added to increase the selectivity. The fundamental problem, which is also common to other types of catalytic sensors, is the catalyst poisoning. In the ThO_2/Al_2O_3 pellistor, the organosilicon, organosulfur, or organophosphorus compounds are among the most serious poisons. A catalyst
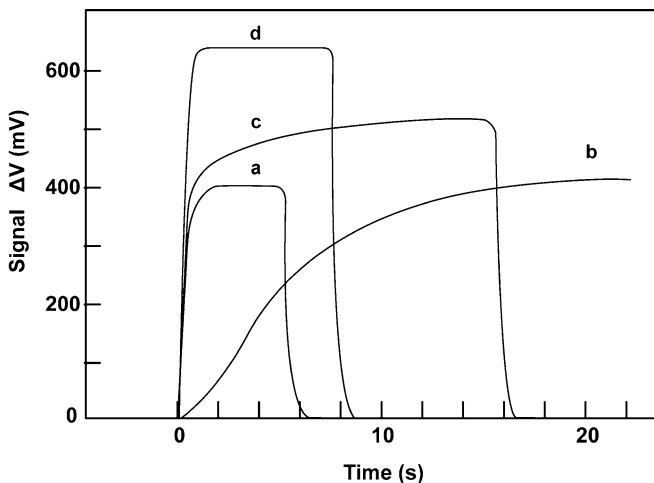
Fig. 3.5 Examples of response of pellistors to (**a–c**) 60% LEL CH$_4$: (**a**) PR-fast response element; (**b**) PR-slow response (diffusion) element; (**c**) conventional Pt/Th pellistor; (**d**) response of PR-fast pellistor to 60% LEL n-butane (adapted from Dabill et al., 1987, p. 135)

consisting of Rh supported on Al$_2$O$_3$ proved to be fast and poison resistant (PR; Fig. 3.5a, b, d; Dabill et al., 1987). Porosity of the catalyst increases the response time (Fig. 3.5c). The coil acts both as the heater and as the resistance thermometer. When the gas reacts at the catalytic surface in a controlled combustion, the evolved heat from the reaction increases the temperature of the pellet and of the platinum coil, thus increasing its resistance. The pellistor is usually connected to one arm of the Wheatstone bridge which provides the output signal. This resistance change can be either detected directly or used in the feedback push–pull operation. For small temperature changes, the out-of-balance signal of the bridge is proportional to the rate of the catalyzed combustion reaction and to the reaction heat ΔH_{rxn}. It is preferable to operate the pellistor in the push–pull mode, at the set temperature of 300–500 °C. A set constant current I_{set} is passed through to the Pt wire at variable voltage, thus adjusting the power output of the electrical heater (3.6). The rate of combustion is described by (3.15).

$$-\frac{dC_G}{dt} = k_r C_G \tag{3.15}$$

Here, k_r is the rate constant and C_G is the concentration of the gas in the porous surface layer of the pellistor. Thus, for detection of CH$_4$, the combustion heat is generated by the reaction

$$CH_4 + O_2 \rightarrow CO_2 + 2H_2O - \Delta H_{rxn} = 890\,kJ/mol \tag{3.16}$$

The rate of heat production causes a proportional decrease of the voltage.

$$\frac{dC_G}{dt}\Delta H_{rxn} \sim -VI_{set} \tag{3.17}$$

Combining (3.15) and (3.17), and including diffusion of the gas to the porous layer yields (3.18).

$$V = \frac{KC_G}{I_{set}}\Delta H_{rxn} \tag{3.18}$$

The proportionality constant K includes the rate constant of the combustion, thermal capacity of the pellistor, and the factor related to the diffusion of the gas. Ideally, all terms on the right-hand side of (3.18) except C_G are constant. For safety applications, the response of the pellistor is expressed on the scale of %LEL, which for methane is %100 LEL $= 5\%$ v/v in air. On the LEL scale, the dynamic range is between 10% and 100%. The response time is around 1 s.
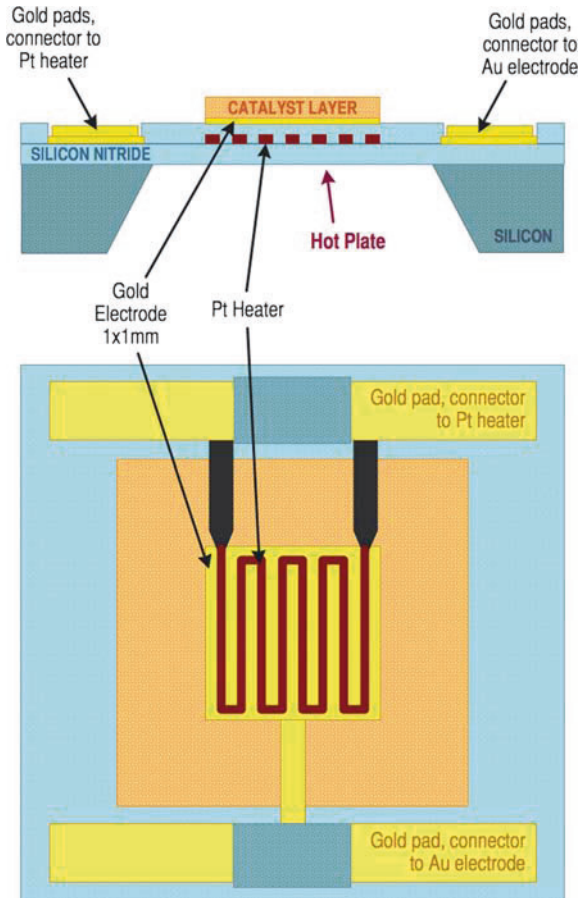
Fig. 3.6 Miniaturized microfabricated pellistor: (a) schematic crossection; (b) the top view (adapted from Bartlett and Guerin, 2003, p. 131)
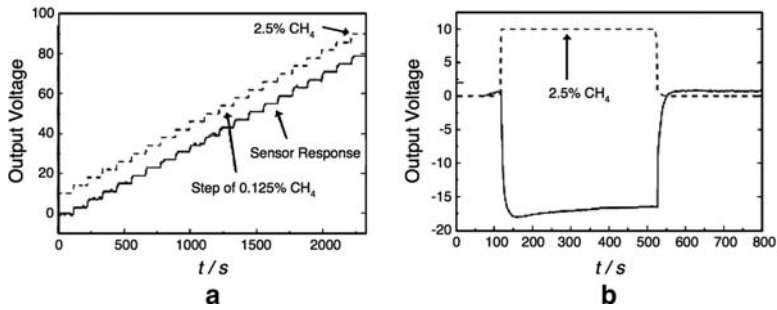
Fig. 3.7 Response of a miniaturized pellistor coated with Pd to methane at 500°C (adapted from Bartlett and Guerin, 2003, p. 128)

3.6.3 Miniaturized Pellistor

The same principles that apply to the original pellistor apply to the microfabricated "planar pellistor" (Fig. 3.6).

Note that the cavity below the Pt heater is etched away, thus significantly reducing the thermal mass. That leads to lower power consumption (175 W at 500°C), faster response time ($t < 60$ s), and higher sensitivity (~35 mV/1% methane). Another significant innovation in this design is the use of a thin gold electrode that allows electrochemical deposition and modification of the selective layer. The response of a miniaturized pellistor electrochemically coated with Pd to methane is shown in Fig. 3.7. The pellistor is operated in a direct mode, using a Wheatstone bridge for detection of resistance changes. However, there is no reason why they could not be operated in the push–pull mode.

Food for Thought #3

Enzyme Thermistor

1. The enzyme thermistors suffer from a notoriously high detection limit when the generated heat is detected as the change of thermistor resistance. Suggest ways of mitigating this problem.

Equivalent Circuit for a Thermal Sensor

2. Using the equivalency relationships in Table 3.1, propose an equivalent electrical circuit diagram for an enzyme thermistor operated (a) in direct detection mode and (b) in the feedback, push–pull mode.

Push–Pull Mode

3. In the push–pull mode, you can adjust the temperature of the heat sink T_2 in any way you want. Is it better to keep the temperature difference $(T_1 - T_2)$ wide or narrow?

Symbols

C_G	Concentration of a gas
C_p	Heat capacity at constant pressure
C_s	Thermal capacity (heat capacity divided by the mass m)
E_g	Band gap energy
ΔH	Change in enthalpy
ΔH_{el}	Heat supplied to the reactor (usually in the form of electric heating)
ΔH_{loss}	Heat loss to the environment
ΔH_{rxn}	Reaction heat
I	Current
K	Proportionality constant
k	Boltzmann constant $(1.38 \times 10^{-23}\,\mathrm{J\,K^{-1}})$
k_r	Rate constant
m	Mass
Q_A	Seebeck coefficient, a material constants of conductor A
R	Resistance
R_T	Resistance at a defined temperature
T	Temperature
V	Voltage

References

Bartlett, P.N. and Guerin, S. (2003) *Anal. Chem.* 75, 126–132.

CRC Handbook of Chemistry and Physics (2006) The Chemical Rubber Company, Cleveland, OH.

Dabill, D.W., Gentry, S.J., and Walsh, P.T. (1987) *Sens. Actuat.* 11, 135.

Herwaarden, A.W. and Meijer, G.C.M. (1994) Thermal sensors. In: S.M. Sze (Ed.) *Semiconductor Sensors*. Wiley, p. 331.

Liess, M. and Steffes, H. (2000) The modulation of thermoelectric power by chemisorption a new detection principle for microchip chemical sensors. *J. Electrochem. Soc.* 147, 3151–3153.

Tran-Minh, C. and Vallin, D. (1978) *Anal. Chem.* 50, 1874.

Chapter 4
Mass Sensors

4.1 Introduction

Change of mass accompanies many interactions of the chemical species with the sensor. Not surprisingly, mass sensors represent an important segment of the chemical sensing field. From the measurement point of view, the determination of mass is called gravimetry. Although scales and balances are standard equipment in any chemical laboratory, they are not usually thought of as sensors. On the other hand, when we talk about microbalances and microgravimetry (Lu and Czaderna, 1984), we regard them as sensors. The transduction principle is the detection of the change of mass through the changes in behavior of some oscillator. Because of their small size, high sensitivity, and stability, piezoelectric crystals have been used as microbalances, namely in the determination of the thicknesses of thin layers and in general gas sorption studies (King, 1964). At the beginning of their development, mass sensors were synonymous with the quartz crystal microbalance (QCM).

Later on, other oscillators, such as the Surface Acoustic Wave (SAW) devices, vibrating beams, and cantilevers, were added. They may employ different oscillatory modes and are made from other, sometimes nonpiezoelectric materials. Incorporation of various chemically sensitive layers has completed the transition from physical microbalance to chemical mass sensor, and has resulted in the increasingly common usage of mass sensors in recent years (Grate et al., 1993). Understanding of the oscillatory mode of operation of these sensors goes beyond detection of mass, and they have become an important tool in the study of interfaces (Janshoff et al., 2000; Schneider and Martin, 1995; Martin and Hager, 1989). This is particularly true for their use in liquids, where the interfacial mechanics often dominates the effects caused by the changes in mass.

The optimum domain of applicability of mass sensors can be evaluated by considering some general aspects of behavior of these devices, from the point of view of the basic species–sensor interaction.

Clearly, a mass-related signal will be obtained only if the species–sensor interaction results in a net change of mass of the chemically selective layer attached to the device. Thus, an equilibrium binding will yield a measurable signal. On the

other hand, if the interaction is just a displacement of one species with another (i.e., exchange or catalytic reaction), the sensor surface is only a temporary host to the interacting species, and the net change in mass may be very small. Therefore, mass sensors rely predominantly on equilibrium, rather than on kinetically based selectivity.

Another limitation comes from the operation of these devices in a liquid phase, as discussed later in greater detail. In such media, the output signal of the sensor is affected not only by the incremental mass change due to the equilibrium binding, but also by other extraneous dynamic parameters of the medium. These include the viscosity of the interfacial region, frictional effects at the sensor–liquid interface (which affect the energy loss), and by changes of solvation at the interface. This is both bad and good news because such secondary changes of the oscillatory behavior can provide valuable information beyond chemical sensing. This is particularly true for most biochemical interactions, where "mass biosensors" elevate these devices beyond their use as purely analytical tools and into the domain of biophysics.

Different types of mass sensors discussed in this chapter are summarized in Fig. 4.1 (Grate et al., 1993). It shows the schematics, principal modes of oscillations, and the common acronyms. Their basic properties and sensitivities are summarized in Tables 4.1 and 4.2, respectively. These mass sensors are based on piezoelectric oscillators. More recent additions to mass sensors are microfabricated cantilevers, which do not rely for their operation on the piezoelectric effect.

The major advantages of mass sensors are the simplicity of their construction and operation, low weight, and small power requirements. In addition, their operating principle depends on a highly reliable phenomenon. The measurement of the frequency shift is one of the simplest and most accurate physical measurements.
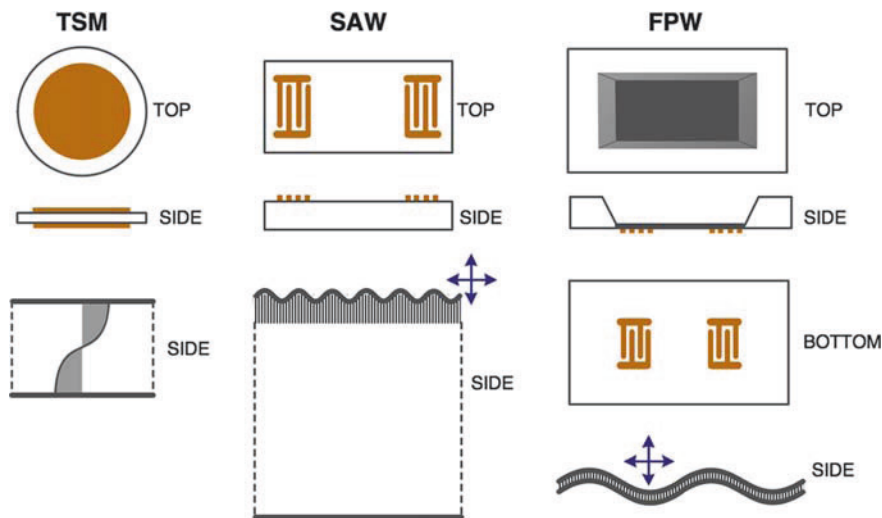
Fig. 4.1 Schematics, modes of vibration, and acronyms of piezoelectric sensors discussed in this chapter (adapted from Grate et al., 1993)

Table 4.1 Comparisons of acoustic waves and devices

Device Type	Wave Type	Particle Displacement Relative to Wave Propagation Direction	Transverse Component Displacement Relative to the Sensing Surface	Media	Plate Thickness	Factors Determining Frequency[a]	Typical Frequency (MHz)	Example
TSM	Bulk	Transverse	Parallel	Gas Liquid	$\lambda/2$	Plate thickness	5 to 10	Quartz 6 MHz $v = 3330\,\mathrm{m/s}$ $d = 277\,\mu\mathrm{m}$ $\lambda = 555\,\mu\mathrm{m}$
SAW	Surface	Transverse, parallel[b]	Normal	Gas	$\gg\lambda$	IDT Finger spacing	30 to 300	Quartz 158 MHz $v = 3160\,\mathrm{m/s}$ $d = 760\,\mu\mathrm{m}$ $\lambda = 20\,\mu\mathrm{m}$
STW	Surface	Transverse	Parallel	Gas[c] Liquid	$\gg\lambda$	IDT Finger spacing[d]	30 to 300	Quartz 250 MHz $v = 5000\,\mathrm{m/s}$ $d = 500\,\mu\mathrm{m}$ $\lambda = 20\,\mu\mathrm{m}$
FPW	Plate	Transverse, parallel[b]	Normal	Gas Liquid	$\ll\lambda$	Plate thickness, IDT finger spacing	2 to 7	Zinc oxide[e] 5.5 MHz $v = 550\,\mathrm{m/s}$ $d = 3.5\,\mu\mathrm{m}$ $\lambda = 100\,\mu\mathrm{m}$
SH-APM	Plate	Transverse	Parallel	Gas[c] Liquid	3 to 10 λ	Plate thickness, IDT finger spacing	25 to 200	Quartz 101 MHz $v = 5060\,\mathrm{m/s}$ $d = 203\,\mu\mathrm{m}$ $\lambda = 50\,\mu\mathrm{m}$

[a]Plate material also plays a role in determining frequency in all cases.
[b]Particle displacement component parallel to the direction of wave propagation is also parallel to the sensing surface. The combination of transverse and parallel components produces a retrograde elliptical particle motion.
[c]Device designed primarily with liquid phase applications in mind, but can be used for gas phase sensing.
[d]With slight modification by the grating.
[e]Composite plate of silicon nitride, aluminum, and zinc oxide: zinc oxide is the piezoelectric material.

Table 4.2 Mass sensitivities of five types of acoustic wave devices

Device Type	Mass sensitivity increases with …	Response to Added Mass[b]	S_m (unperturbed)[c]	S_m Example (Hz/MHz)/(ng/cm²)[d]
TSM	Decreasing plate thickness (Increasing frequency)	$\Delta F = -C_T F^2 \Delta M$	$S_m = -\frac{2}{\rho\lambda} - \frac{1}{\rho t}$	0.014 (6 MHz quartz plate)
SAW	Decreasing wavelength (Increasing frequency)	$\Delta F = -C_S F^2 \Delta M$	$S_m = \frac{K(\sigma)}{\rho\lambda} = -\frac{K(\sigma)}{\rho V_R} F$	0.20 (158 MHz quartz plate)
STW	Decreasing wavelength (Increasing frequency) quartz plate Increasing grating thickness	$\Delta F = -C_G F^2 \Delta M$	$S_m = -\frac{K'(\sigma,G)}{\rho\lambda} = -\frac{K'(\sigma,G)}{\rho V_{STW}} F$	0.180 (250 MHz quartz plate)
FPW[e]	Decreasing plate thickness (Decreasing frequency for constant wavelength)	$\Delta F = -\frac{1}{2M} F \Delta M$	$S_m = -\frac{1}{2M} - \frac{1}{2\rho t} =$	0.38 (5.5 MHz composite plate)
SH-APM[f]	Decreasing plate thickness (Increasing frequency for constant wavelength)	$\Delta F = -\frac{J}{\rho t} F \Delta M$	$S_m = -\frac{J}{\rho t}$	0.019 (101 MHz quartz plate)

[a]Symbols: F, device frequency; ΔF, frequency shift; ΔM, change in mass per unit area; C_T, a constant dependent on the properties of the piezoelectric plate of the TSM device; C_S, a constant dependent on the properties of the piezoelectric plate of the SAW device; C_G, a constant dependent on the properties of the piezoelectric plate and grating of the STW device; $K(\sigma)$, a factor dependent on the properties of the piezoelectric plate of the SAW device; $K'(\sigma,G)$, a factor dependent on the properties of the piezoelectric plate and grating of the STW device; M, the mass per unit area of the FPW membrane; ρ, the density of the plate material; d, the plate thickness; λ, the acoustic wavelength; V_R, the Rayleigh wave velocity; V_{STW}, the shear transverse wave velocity; S_m, defined in the text. (Grate et al. 1993)

[b]Assumed to be evenly distributed.

[c]Bare device in gas or vacuum.

[d](Hz/MHz)/(ng/cm²) can be converted to cm²/g (units often used in other publications) by multiplying by 1000.

[e]FPW composite membrane has zinc oxide piezoelectric layer. Second expression for FPW S_m is for the special case of single layer isotropic membrane.

[f] SH-APM expressions are approximations for an isotropic plate, and are for individual modes, where J = 1/2 for n=0, and J = 1 for n>0. In the example, 101 MHz is the frequency for the n=0 mode, while the mass sensitivity is for the n=1 mode.

Moreover, impedance analysis provides a powerful approach to the study of interfaces. Finally, mass sensors have high sensitivity, and can be used with a variety of selective layers for sensing of a very broad range of compounds.

One pitfall to watch for is their relatively high temperature sensitivity. Incorrect interpretation of frequency shifts as being due to mass changes alone can also yield flawed results. This can be particularly troublesome with mass biosensors. Proper experimental design can mitigate the temperature effects, however, knowledge of the underlying physics and rigorous verification of the source of the signal variation are the surest protection against misinterpretation of the raw data.

4.2 Mass Sensors Based on Piezoelectric Materials

There are several types of materials that exhibit the piezoelectric effect. Because it is inexpensive, and because it has a relatively strong piezoelectric coefficient, quartz is the material of choice for most piezoelectric sensor applications. It has a hexagonal crystallographic structure, with no center of symmetry. Both the magnitude of the piezoelectric coefficient and the extent of its temperature dependence are affected by the orientation of the cut of the crystal with respect to the main crystallographic axes. The most popular AT-cut is shown in Fig. 4.2.

For practical reasons, the optimum orientation is chosen such that the crystal has a minimum temperature dependence within the intended operating temperature range. Most applications use the AT-cut $35°15''$ inclined in the yz-plane, which exhibits a minimum temperature coefficient between 50–70$°$C (Janshoff et al., 2000). Yet at the ambient temperatures (25–45 $°$C) encountered in sensing
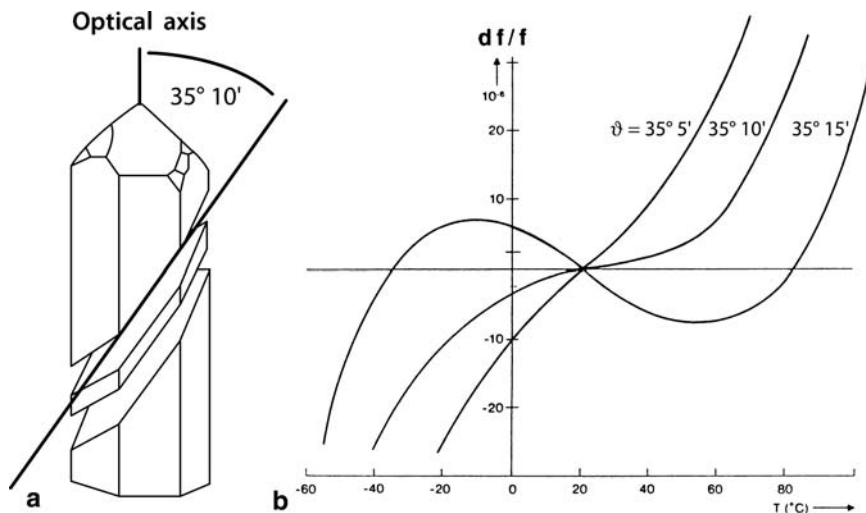

Fig. 4.2 (**a**) Orientation of the AT-cut of the quartz crystal; (**b**) Dependence of relative frequency of quartz crystal microbalance (QCM) on temperature

applications, the $35°10''$ would be clearly preferable. The reason for this apparent discrepancy lies in the fact that the main application of quartz crystals is not in chemical sensing, but in electronics, where the ambient operating temperature inside the electronic instrument is higher. As we see time and time again, chemical sensors have been the byproduct of technologies that were developed for applications other than chemical sensing. Another example is optical fibers, which were developed for communications, but now find use as optical sensors (Chapter 9) or transistors for general electronics.

Piezoelectricity is the effect that links mechanical stress to the displacement of charge, leading, in turn, to an electric field. Thus, mechanical force applied to piezoelectric material induces an electric field in the crystal. Because charge leaks, application of a steady force produces only transient voltage, which can be as high as 20,000 V. For this reason piezoelectricity is almost exclusively studied and used in AC mode. This fact is, however, very beneficial from the measuring point of view because, as mentioned previously, frequency is one of the most precisely measurable quantities. In fact, precision approaches 1:1000 and remains stable over 1 month. The quantitative relationship governing the operation of these sensors therefore links the mass with the change of frequency.

4.2.1 Thickness-Shear-Mode (TSM) Resonators

This is the correct name for most popular mass sensors, although they are better known as Quartz Crystal Microbalances (QCMs). A piezoelectric crystal vibrating in its resonance mode is a harmonic oscillator. For microgravimetric applications, it is necessary to develop quantitative relationships between the relative shift of the resonant frequency and the added mass. In the following derivation, the added mass is treated as "added thickness" of the oscillator, which makes the derivation more intuitively accessible.

Figure 4.3 shows the shear-mode vibration of a quartz crystal of mass M and thickness t. At resonance, the wavelength λ is

$$\lambda = 2t/n \tag{4.1}$$

where $n = 1$ designates the frequency F and $n > 1$ its higher harmonics. The shear velocity v is defined as

$$v = \lambda F \tag{4.2}$$

Therefore,

$$F = \frac{v}{2t} \tag{4.3}$$

The change of resonance frequency, caused by the infinitesimally small change of the crystal thickness dt, is then obtained by differentiating (4.3).

$$dF = -\frac{v}{2t^2}dt \tag{4.4}$$
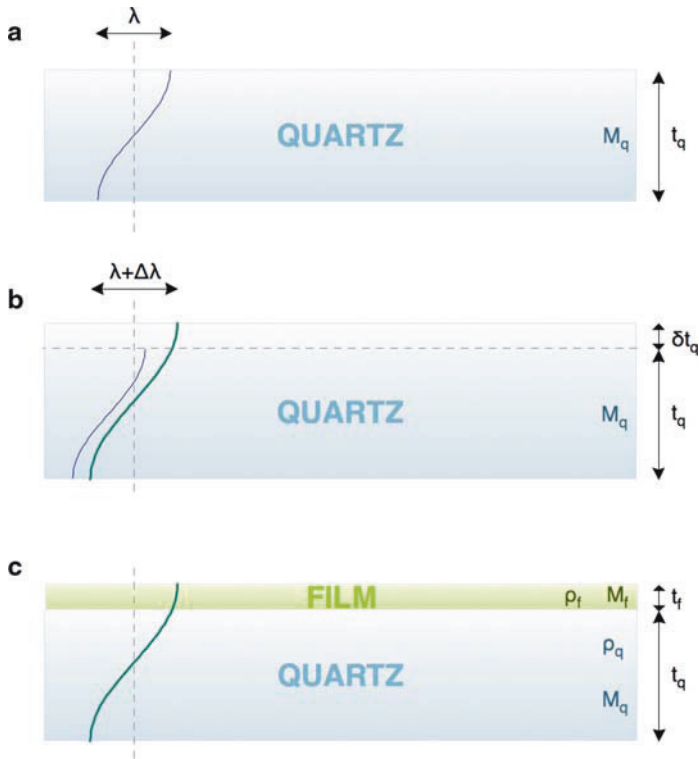
Fig. 4.3 Shear mode vibrations in a thickness-shear-mode (TSM) device. Incremental increase $\Delta\lambda$ is due to increased thickness

Dividing (4.4) by (4.3) yields the relative change of frequency.

$$\frac{\mathrm{d}F}{F} = -\frac{\mathrm{d}t}{t} \tag{4.5}$$

Thus, the relative increase of crystal thickness (shown in Fig. 4.3) lowers the resonant frequency. Equation (4.5) can be written in terms of the total mass M deposited per unit area and its finite change ΔM as

$$\frac{\Delta F}{F} = -\frac{\Delta M}{M} \tag{4.6}$$

In this derivation, the increment of the foreign mass ΔM is assumed to be uniformly distributed as a thin film over the crystal surface. It is treated as equivalent to a change in mass of the whole crystal. Therefore, the approximation $\Delta M_{\text{crystal}} \sim \Delta M_f$ is made. It means that the added mass behaves as a rigid film having the same properties, namely the shear velocity, as the quartz crystal. With this in mind, we use the subscript "f" to remind us that this film is a "foreign added mass," whereas symbols referring to the crystal itself are without subscript. The relative change of

resonant frequency due to the film of density ρ_f and thickness t_f deposited over a unit area is

$$\frac{\Delta F}{F} = -\frac{\rho_f t_f}{\rho t} \tag{4.7}$$

Substitution for crystal thickness t from (4.3) yields the relative change of the resonant frequency.

$$\frac{\Delta F}{F} = -\frac{\rho_f t_f 2F}{v\rho} \tag{4.8}$$

Alternatively,

$$t_f = -\frac{\Delta F}{2F^2}\frac{v\rho}{\rho_f} \tag{4.9}$$

Constants v, ρ, and F are the material properties of the crystal. Because the density of the deposited film, ρ_f, is known, the QCM becomes a thickness monitor. (In fact, monitoring of thickness of thin films during their deposition was the original application of QCM.)

From (4.8) we can define the crystal constant C_q as

$$C_q = \frac{2}{v\rho} \tag{4.10}$$

Because $\rho_f t_f = \Delta M_f$, the relationship between frequency shift and added mass M_f becomes

$$\Delta F = -C_q F^2 \Delta M_f \tag{4.11a}$$

The relative change is indicated by

$$\frac{\Delta F}{F} = -C_q F \Delta M_f \tag{4.11b}$$

This relationship is known as the Sauerbrey equation; it is the basic transduction relationship of the QCM when it is used as a chemical sensor. Due to the assumptions made throughout this derivation, the Sauerbrey equation is only semi-quantitative. The assumption of the added "rigid mass" mentioned earlier is its most serious limitation. The material added to the QCM will invariably exhibit different mechanical characteristics than quartz itself. Thus, the assumption of unified behavior is weak at best.

Other factors, such as mechanical clamping, damping in the electrical circuit, and temperature also affect the absolute accuracy. For this reason it is necessary to use calibration curves for quantitative work. In spite of these limitations, the quartz microbalance is an extremely sensitive and versatile sensor.

The "Sauerbrey limit" is often cited as the upper amount of added mass beyond which the equations (4.11a, b) do not hold. It is different for different materials. For an AT-cut quartz ($\rho = 2,650\,\text{kg}\,\text{m}^{-3}$; $v = 3,340\,\text{m}\,\text{s}^{-1}$) which has a resonant frequency at $F = 5\,\text{MHz}$, the crystal constant is $C_q = 5.65\,\text{MHz}\,\text{cm}^2\,\text{kg}^{-1}$. This means that a 1 Hz shift corresponds to $17.7\,\text{ng}\,\text{cm}^{-2}$ of added weight. The dynamic range extends up to $20\,\mu\text{g}\,\text{cm}^{-2}$, which often exceeds the Sauerbrey limit.

4.2.1.1 QCM in the Gas Phase

Oscillation of the crystal is due to the AC electric field applied across the crystal from metal electrodes. Typical electric fields are quite low, 10–20 V cm^{-1}. The electrodes are deposited directly on the quartz crystal (Fig. 4.4) although, in principle, a noncontact activation of the crystal is also possible.

One of the electrodes is smaller in order to eliminate the "edge effect," that is, fringing electric field. Also, the metal is deposited in a "keyhole" pattern, with each electrode rotated 180° with respect to the other (Janshoff et al., 2000). This geometry is needed in order to make the field as cylindrical as possible. Commercial crystals have a "flat," which marks the orientation of the crystal with respect to the crystallographic coordinates. The fundamental frequency of the crystal depends on its thickness (4.3). Thus, a 5 MHz crystal is 330 μm thick and a 30 MHz crystal has a thickness of 55 μm. The thickness of the metal electrode also plays an important role. There is no piezoelectric displacement outside the electric field. This means that, at the boundary between the metal and the metal-free region, the energy is reflected, resulting in the formation of a standing wave in the excited part of the crystal, beneath the metal electrodes. This effect is called energy trapping. It occurs when the metal is at least 50 nm thick. On the other hand, if the metal is too thick, its mass exceeds the Sauerbrey limit, and the crystal ceases to oscillate.

There are two modes excited by the AC field, longitudinal and transverse. For crystals in the 100–300 μm thickness range, only the transverse standing wave needs to be considered (Janshoff et al., 2000). The actual lateral displacement of a point on the crystal surface (and therefore the mass sensitivity) is the Gaussian function of the radial distance from the center of the electrode (Fig. 4.5). It also depends on the amplitude of the applied electric field and ranges from few nm/V in water to tens of nm/V in air or in vacuum.

The direction of the displacement comes from the periodic boundary condition of the solution of the wave equation. It is normal to the "flat" of the crystal
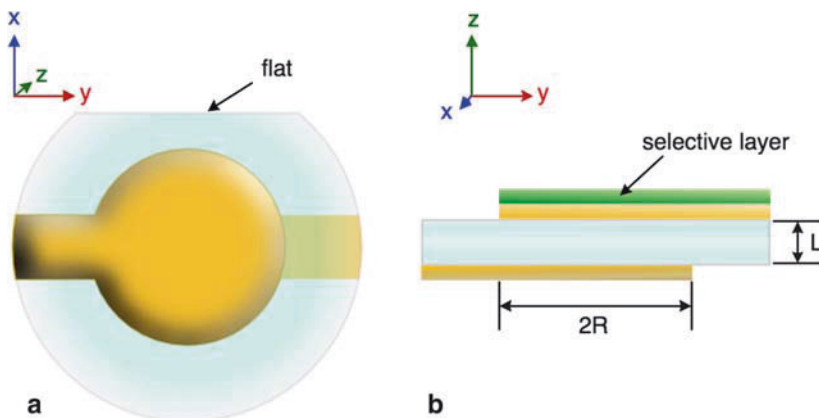
Fig. 4.4 Practical quartz crystal microbalance (QCM) with "keyhole" electrodes and "flat"
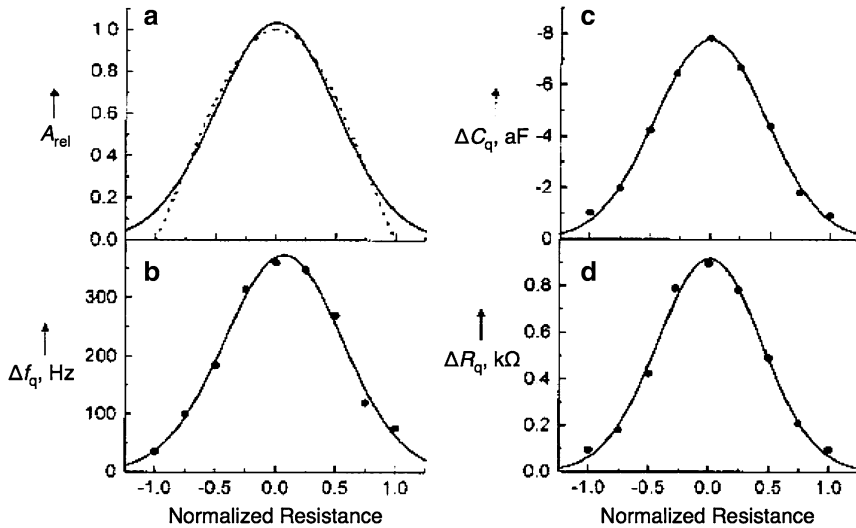
Fig. 4.5 Radial distribution in quartz crystal microbalance (QCM) of (**a**) displacement amplitude; (**b**) mass sensitivity; (**c**) elasticity; and (**d**) frictional losses (adapted from Janshoff et al., 2000, F9)
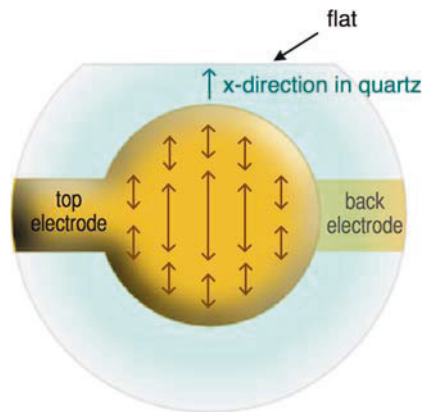
Fig. 4.6 Surface velocities of the displacement vector indicating radial distribution and orientation in the *x*-direction (adapted from Martin and Hager, 1989a)

(Fig. 4.6; Martin and Hager, 1989a, b; Hillier and Ward, 1992). The above considerations make the description of the physics of the QCM far more complicated than is apparent from the simple Sauerbrey equation.

Another level of complexity must be considered: the oscillations in the *z*-direction. For chemical sensing, we need to impart the selectivity by depositing an additional selective layer on top of the smaller of the two electrodes. Therefore, on a quartz crystal that is used for chemical sensing, we have at least three interfaces: one between the quartz and the metal, a second one between the metal and the selective layer, and the third between the selective layer and the sample (Fig. 4.4b).

At each interface, reflection and refraction of the acoustic energy occur similar to the reflection and refraction of light at the boundary between two materials of different optical densities (see Chapter 9).

Let us first consider the quartz–metal interface. In deriving the Sauerbrey equation, a pivotal assumption was made, namely that the shear velocities in the crystal and in the film are equal. By this assumption we perform a virtual transformation from "thickness" to "mass." This is the weakest point of Sauerbrey derivation. Here we do not make the same assumption. The shear velocities in the crystal and in the film are

$$v = \left(\frac{\mu}{\rho} \right)^{1/2} \tag{4.12a}$$

and

$$v_f = \left(\frac{\mu_f}{\rho_f} \right)^{1/2} \tag{4.12b}$$

Here μ is the shear modulus of the respective layer. Let us also assume that there is no frictional loss, that is, that the resonance condition applies. Then

$$\tan \left(\frac{\pi F_f}{F} \right) = - (\rho_f v_f / \rho v) \tan \left(\frac{\pi \Delta F_f}{F} \right) = - \frac{\rho_f v_f}{\rho v} \tag{4.13}$$

where $F_f = \omega/2\pi$ is the resonant frequency of the crystal with material deposited on it. The terms $\rho_f v_f$ and ρv are the acoustic impedances Z_f and Z of the film and the crystal, respectively. Their ratio ($\zeta = Z_f/Z$) is an important parameter with respect to the acoustic matching of the two materials. For optimum resonant conditions, ζ should be as close to unity as possible. Different materials and corresponding values of Z are listed in Table 4.3. Using the values in the table, we can see that the quartz–Al interface is a good match ($\zeta = 0.99$) and that for the quartz–Au interface, ζ has

Table 4.3 Selected values of acoustic impedances of materials used as electrodes in mass sensors

Material	Shear Mode Acoustic Impedance $(10^6 \, \text{kg s}^{-1} \, \text{m}^{-2})$
Aluminum	8.22
Aluminum oxide	24.6
Chromium	29.0
Copper	20.3
Gold	23.2
Graphite	2.71
Indium	10.5
Nickel	26.7
Palladium	24.7
Platinum	36.1
Quartz	8.27
Silicon	12.4
Silver	16.7

an acceptable value at 2.80. On the other hand, QCMs with Pt electrode (for quartz–Pt, $\zeta = 4.36$) result in a low Q-factor, which often does not oscillate.[1] The choice of the metal is clearly dictated by the application. In electrochemical experiments, for example, gold or platinum are often the only practical materials, although their Q-factor is not entirely optimal.

When selective layers are deposited, the whole structure must be treated as a multiple resonator in which the reflection and/or refraction of the acoustic energy occurs at each interface. For example, when a polymer film is deposited on top of the gold electrode of the QCM, it is the polymer–Au interface with which we are concerned. When the mass loading of multiple structures becomes too high, the effect of the impedance mismatches becomes significant and the crystal ceases to oscillate. Even approximate treatment of the multiple resonator is difficult because densities, as well as thicknesses and shear moduli, of the individual layers must be known.

Other, secondary effects skew the response of the piezoelectric sensor when it is used as a gas sensor. These can be described phenomenologically as the hydrostatic effect (p), the frictional effect (x), and the sorption effect (m):

$$-\frac{\Delta F}{F} = \left(\frac{\Delta F}{F}\right)_p + \left(\frac{\Delta F}{F}\right)_x + \left(\frac{\Delta F}{F}\right)_m \qquad (4.14)$$

It is, of course, the last effect – selective adsorption (or absorption) of the species of interest – which is the key for chemical sensing; the first two effects can be regarded as nonspecific interferences. There are, however, situations in which these two effects cannot be ignored, even in gas phase applications. Thus, for an AT-cut crystal at $50\,^\circ$C, (4.14) is

$$-\left(\frac{\Delta F}{F}\right) \cdot 10^6 = 1.35 10^{-3} P_{\text{torr}} + 7.2 10^{-3} (\pi F \rho_{\text{g}} \eta_{\text{g}})^{1/2} + 2.26 F \Delta M \qquad (4.15)$$

where P is the gas pressure, ρ_{g} and η_{g} are the gas density and viscosity, respectively, M is the mass in grams per unit area, and F is the crystal frequency. Thus, for a 5 MHz AT-cut quartz crystal, the change of frequency upon transition from air at 1 atm to vacuum ($\sim 10^{-6}$ mmHg) is -5 Hz. This change is due only to the hydrostatic effect and to the increased friction of the crystal experiences in air.

4.2.1.2 QCM in Aerosols and Suspensions[2]

The QCM can be used for monitoring heterogeneous samples, such as aerosols and suspensions. The mass increase due to impacting and sticking particles

---

[1] "Quality factor" (Q-factor) compares the frequency of oscillation to the rate of dissipation of energy of the oscillating system. Higher Q indicates less energy dissipation, relative to the oscillating frequency.

[2] Ho (1984).

(liquid–aerosol, or solid–suspension) can be treated by the Sauerbrey equation. In that case, the particle diameter plays an important role. Increased diameter of the particle increases the contact area, but on the other hand, also increases the inertial force which tries to dislodge the particle from the surface. Thus, the optimum diameter (D_{crit}) of the particle, which is inversely proportional to the frequency of the crystal, can be expressed empirically as

$$D_{crit} = \frac{\beta}{F_0} \tag{4.16}$$

The parameters included in β are the particle density, the adhesion, and the power dissipated by the crystal. Thus, the variation of frequency provides some means for discrimination based on the particle size. The sensitivity decreases for particles with diameter greater than $2\,\mu m$ and no sensitivity is obtained for particles larger than $D \sim 20\,\mu m$.

Because the mass sensitivity of the QCM is a function of radial distance, being highest in the center (Fig. 4.5b), the lateral particle distribution also affects the magnitude of the response. The "stickiness" of the crystal can be enhanced either chemically or electrostatically. Figure 4.7 shows the schematic of a QCM-based instrument for detection of aerosols in which the crystal is one of the electrodes to which a DC voltage is applied. In order to chemically increase adhesion, coatings of polysiloxanes or silicone grease have been used.

The concentration of aerosol is usually expressed as "mass of liquid phase/sample volume." The collection efficiency (E_c) of the sensor is then defined as

$$E_c = \Delta M / M \tag{4.17}$$

where ΔM is the mass deposited on the crystal. Assuming $E_c = 1$, a mass detection limit of $1\,\mu g\,m^{-3}$ of air sample can be obtained.
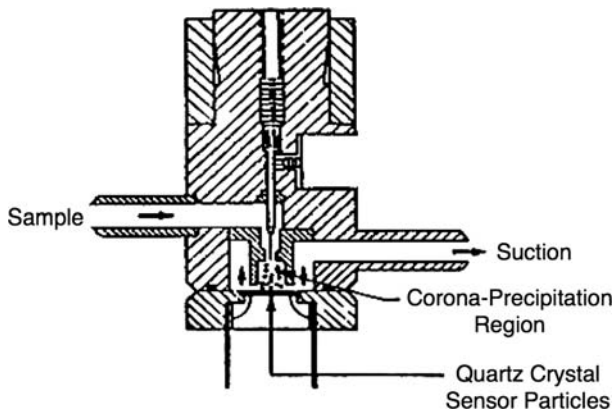

Fig. 4.7 Quartz crystal microbalance (QCM)-based instrument for aerosol analysis (adapted from Ho, 1984)

4.2.1.3 QCM in Liquids

The success of QCM in gas-phase sensing applications has created interest in
its application in liquid media, especially in water. The general idea has been to
combine the often exceptional selectivity of biochemical reactions (e.g., immuno-
chemical or DNA recognition) and to use mass sensing as an apparently simple
transduction principle (Bizet et al., 1998; Marx, 2003). However, it has been
suggested (Yang and Thompson, 1993) that as the chemical nature of the interface
changes, the frictional effects may overshadow any mass effects. On the other hand,
if the chemical composition of the interface does not change, the frictional losses at
the interface remain constant and the resonant frequency of the crystal follows the
Sauerbrey equation within its normal limits. This is the case in the electroplating
of metals onto the QCM electrode, or in detecting the increasing concentration of
glucose at the electrode. Unfortunately, such applications cannot be regarded as true
chemical sensing, because the chemical composition of the sample does not change.

Operation of the QCM in liquids increases the complexity to yet another level. It
extends the application of QCM beyond chemical sensing and turns it into a power-
ful tool of interfacial studies. Let us first outline the physics of this operation. First
of all, the two electrodes cannot be in the same conducting electrolyte because they
would form an electrochemical cell and short out. Therefore, only the side of the
QCM with the smaller electrode is usually exposed to the liquid. If both electrodes
are exposed to the solution, the two compartments must be electrically isolated.

One of the first and most elegant applications of the quartz microbalance was
monitoring of thickness during electroplating of metals. Because liquid represents
an additional mass load, the coupling of the transverse elastic shear wave to the
fluid has to be taken into account. In a liquid medium, the shear motion of the crys-
tal causes the adjacent layer of the solvent (water) to move along the surface (Martin
and Hager, 1989a). The next, more loosely bound, layer of water moves with a cer-
tain amount of "slip," which corresponds to a phase shift for the third and subsequent
layers. Thus, a wave in the perpendicular (normal) direction is established, which
dissipates energy to the liquid and results in acoustic attenuation. It is attributable to
viscous loss which, for Newtonian fluids, results in a damped wave. It extends to the
solution over exponentially decreasing damping distance δ (Kanazawa and Gordon,
1985). The magnitude is typically a fraction of a μm; for example, $\delta \sim 250\,\text{nm}$ for a
5 MHz crystal in water.

$$\delta = \left(\frac{\kappa_L}{\pi F} \right)^{1/2} \tag{4.18}$$

It increases with increasing kinematic viscosity ($\kappa_L = \eta_L / \rho_L$) of the liquid
(Fig. 4.8).

For a Newtonian liquid, the adhering layer moves in phase with the oscillating
crystal and behaves as a "rigid mass." In that case, the damping distance can be
substituted for film thickness t, and the mass loading can be expressed as

$$\Delta M = \rho_L \delta \tag{4.19}$$
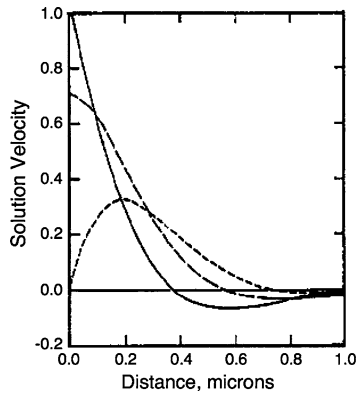
Fig. 4.8 Dissipation of shear wave velocity with distance (adapted from Kanazawa and Gordon, 1985)
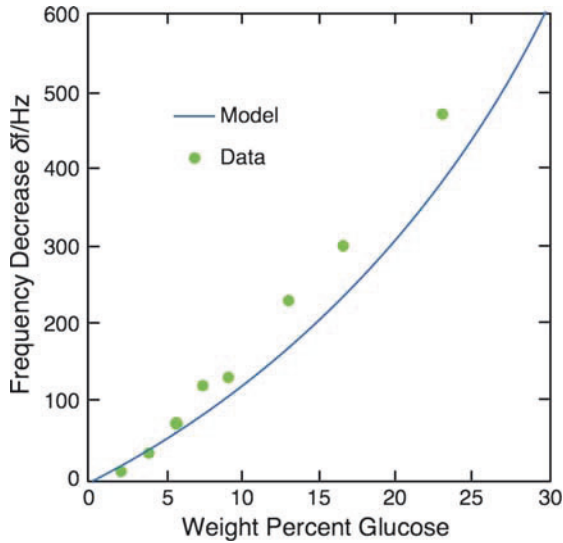
Fig. 4.9 Effect of kinematic viscosity on resonant frequency (adapted from Kanazawa et al., 1985)

The corresponding frequency shift is (Kanazawa and Gordon, 1985)

$$\Delta F = -F_0^{3/2} \left(\frac{\rho_L \eta_L}{\pi \mu \rho} \right)^{1/2} \tag{4.20}$$

The effect of the microscopic kinematic viscosity at the crystal–liquid interface is particularly important. This can be seen in Fig. 4.9, where the discrepancy between predicted and measured frequency is apparent. The discrepancy is not related to the added mass of the glucose; the model accounts for that difference. The deviation
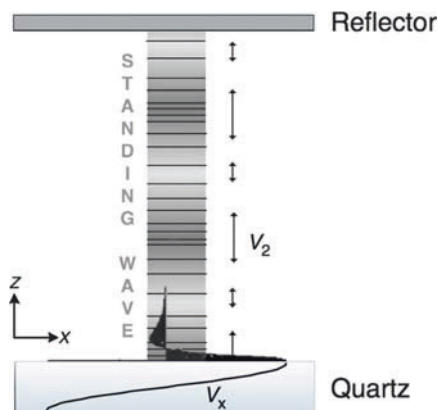
Fig. 4.10 Compressional standing wave created by proximity of the reflecting wall (adapted from Janshoff et al., 2000)

arises from the increased density of the solution, with increased concentration of glucose.

Thus, the effect of "microviscosity" in condensed media has an effect on frequency that can be comparable to or larger than the frequency shift caused by the addition of mass. This interference has the same origin as the frictional interference observed with the oscillating crystal in gas (4.15), except that it is much more severe in liquids.

Another interesting effect that plays a role in the design of the proper housing for QCM in liquid is the generation of compressional standing waves. This can be observed when the crystal oscillates in parallel with a vessel wall in close proximity, which acts as a reflector (Fig. 4.10; Janshoff et al., 2000; Schneider and Martin, 1995). The effect can be observed over the distance of the reflector surface (e.g., 100 µm; Lin and Ward, 1995; Schneider and Martin, 1995).

Trapping of liquid in the rough surface of the electrode adds "virtual mass" and may also cause an additional mass loading artifact (Theisen et al., 2004). This effect can be particularly severe when porous materials such as conducting polymers are deposited at the QCM electrode.

In summary, the parameters that affect the performance of the QCM as a chemical sensor in liquids include excessive mass loading due to the mass of the chemically selective layer itself, the change of mechanical properties of the selective layer, the stress caused by the electrode mounting, positioning of the QCM inside the fluid compartment, conductivity of the liquid, and porosity leading to liquid trapping. Some of these problems can be mitigated by the use of a reference crystal (differential measurement; Bruckenstein et al., 1994; Dunham et al., 1995). Another possibility is to try to understand and use their origin for gaining additional knowledge about the state of the sensor–sample interface. A powerful tool used in this approach is the modeling of the system by equivalent electrical circuits, as discussed below.

Modeling with Equivalent Electrical Circuits

This modeling approach can be applied to different types of chemical sensors, particularly to the study of their dynamic behavior. We have seen the first hint of this approach in Thermal Sensors (Table 3.1). It is related to the operations performed by now-largely extinct analog computers, which were well suited for solving complex systems of higher order and partial differential equations.

There are multiple steps in performing this analysis. First, the elements of physical processes are identified, such as energy storage, energy dissipation, mass transport, and the like. In the second step, the electrical elements corresponding to these processes are identified by their function. For instance, a capacitor can be assigned to represent energy storage or charge separation. A simple resistor can be identified with hindered flow of heat or fluid. The so-called transmission line, consisting of a distributed network of capacitors and resistors, is used to model some nonlinear distributed parameters. In some cases, a diode can represent some unidirectional nonlinear process, and so on.

In the next step, these equivalent electrical circuit elements are arranged (i.e., electrically connected) in a model electrical circuit. This step is both visual and intuitive. Next, the excitation signal is provided corresponding to the dynamics of the problem. For a thermal sensor, this signal is a DC voltage; in the modeling of dynamic systems, such as thickness-shear-mode resonators or in some electrochemical problems, we may use an AC signal. The angular frequency ω then becomes another important modeling parameter.

Finally, the circuit is "solved." It can be done in two ways. Either a corresponding physical circuit is actually built from these electrical components and its transfer function (H) is measured, or the transfer functions of the individual elements and of the entire circuit are represented by explicit mathematical equations and the transfer function is calculated. There are always certain eigenvalues in these solutions that aid in the assignment of physical meaning to the calculated parameters. The transfer function is defined as the frequency-dependent ratio of the output voltage to the input voltage. It is a complex variable for an AC excitation signal.

$$H = \frac{V_{\text{out}}(\omega)}{V_{\text{in}}(\omega)} \tag{4.21}$$

The output of the model is then compared with the output of the real device and the individual elements are iteratively adjusted. When a good fit is obtained, the model is tested. It is a very important step, because the robustness of this procedure must be characterized by establishing the range of validity of the model, for the frequency and amplitude of the excitation signal, as well as for the range of values of the individual circuit elements. The wider the validity range, the more accurate is the representation of the real device by its model. The flowchart for building the equivalent electrical circuit model is shown in Fig. 4.11, and the equivalent electrical circuit of a QCM harmonic oscillator is shown in Fig. 4.12. Close to its resonance,
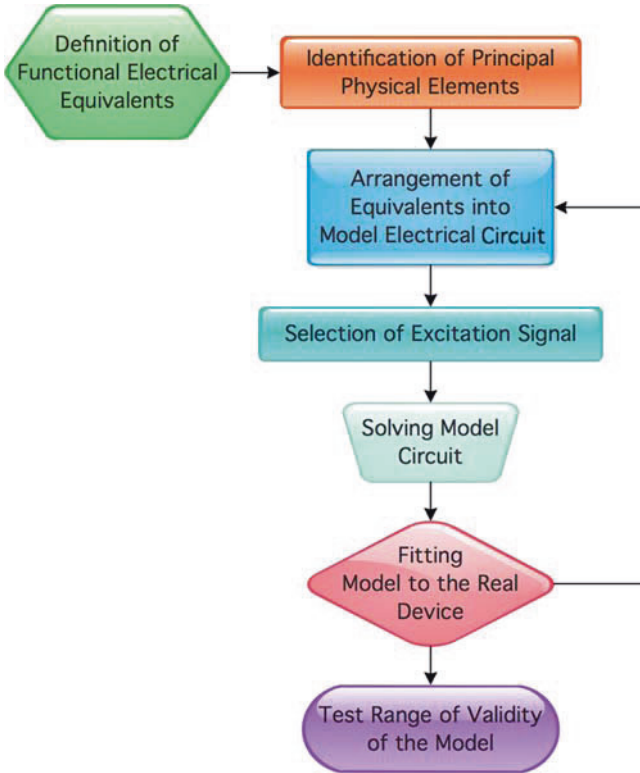
Fig. 4.11 Flowchart for equivalent electrical circuit modeling
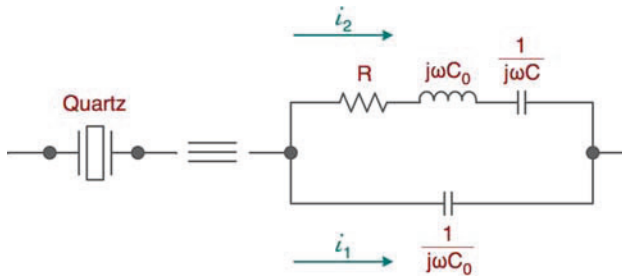
Fig. 4.12 Butterworth–van Dyke (BVD) equivalent circuit

this model is known as a Butterworth–van Dyke (BVD) circuit which was proposed for broadband frequency operation of the QCM by Mason (1965).

In this model, it is assumed that the energy of the oscillator is stored in a capacitor **C**, its losses are represented by resistor **R** and the mass loading by inductor **L**. The "static" capacitance of the crystal with the electrodes is shown as capacitor C_0. The transfer function of this oscillator is governed by two equations.The first is
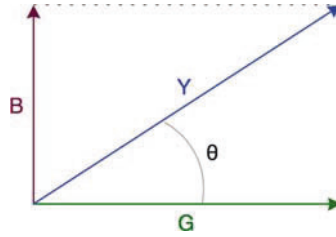
Fig. 4.13 Definition of complex plane numbers

the generalized Ohm's law (4.22).

$$\frac{1}{Y(\omega)} = Z(\omega) = \frac{V(\omega)}{I(\omega)} \tag{4.22}$$

The second is the generalized Kirchhoff's law (4.23) for the current divider.

$$I(\omega) = i_1(\omega) + i_2(\omega) \tag{4.23}$$

The variables are frequency-dependent and represent Y (admittance), Z (impedance), V (voltage), and I, or i (current). The relationship between angular (ω) and linear (F) frequency is $\omega = 2\pi F$. Both the admittance and the impedance are complex numbers, consisting of real and imaginary parts. Thus for admittance

$$Y(\omega) = G(\omega) + jB(\omega) \tag{4.24}$$

where $j = (-1)^{1/2}$, G is called conductance (the real part), and B is called susceptance (the imaginary part). Their relationship in the complex plane (Fig. 4.13) defines the phase angle θ for which

$$\tan\theta = \frac{B}{G} \tag{4.25a}$$

and

$$\cos\theta = \frac{G}{Y} \tag{4.25b}$$

Our task is now to figure out the current flow according to (4.23) through the BVD circuit, respectively through its two arms at different frequencies. The instruments used to do this job are network analyzers (Fig. 4.14), which measure admittance Y (or impedance Z) and the phase angle θ, for different frequencies, in so-called Bode plots. As an alternative, one can separate the real and the imaginary parts of the impedance and present them as so-called Nyquist (or Johnson) plots. In both cases, the objective is to identify the eigenfrequencies at which equations simplify to the extent that the measured values can be converted to separated values of C, R, and L with their equivalent physical meaning.
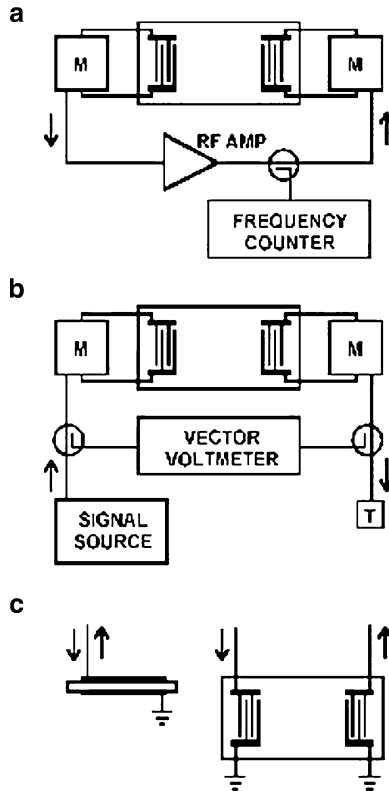
Fig. 4.14 Instruments for network analysis (adapted from Grate et al., 1993)

We first follow the route leading to the Nyquist plot. In the BVD circuit the current passing from input to output encounters the following impedances.

$$\text{For resistor } \mathbf{R}: \; Z = R \; Y = 1/R \qquad (4.26a)$$

$$\text{For inductor } \mathbf{L}: \; Z = j\omega L \; Y = 1/j\omega L \qquad (4.26b)$$

$$\text{For capacitor } \mathbf{C}: \; Z = 1/j\omega C \; Y = j\omega C \qquad (4.26c)$$

The overall admittance of the QCM oscillator is then

$$Y = j\omega C_0 + \frac{1}{R + j\omega L + 1/j\omega C} \qquad (4.27)$$

The first term in (4.27) represents the admittance of the "stationary arm" through which current i_1 flows. The second term is the admittance of the "motional arm" and it represents the path for current i_2. Next, we perform the separation of the real and imaginary parts according to (4.24) (see Appendix D).

$$G = \frac{R}{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} = Y_{\text{Re}} \tag{4.28}$$

and

$$B = \omega C_0 - \frac{\left[\omega L - (\omega C)^{-1}\right]}{R^2 + \left[\omega L - (\omega C)^{-1}\right]^2} = Y_{\text{Im}} \tag{4.29}$$

Solving for ω yields equation of circle of radius $1/2\,R$.

$$\left(\frac{1}{2R}\right)^2 = \left(G - \frac{1}{2R}\right)^2 + (B - \omega_s C_0)^2 \tag{4.30}$$

In this equation, ω_s is the "stationary" frequency corresponding to the zero admittance of the stationary arm. For $G = 0$, $B = \omega_s C_0$ and for $B = \omega_s C_0$, $G = 1/R$ (or zero). These conditions define the position of the center of the circle in the G–B coordinates, at $G = 1/2R$ and $B = \omega_s C_0$ (Fig. 4.15). In this admittance Nyquist plot, the frequency ω increases in the clockwise direction. (Note: In the equivalent impedance Nyquist plot, the frequency increases in the counterclockwise direction.)

On this plot, we can identify some important points. The horizontal line through the center intersects the circle at points **O** and **R** that represent the frequencies
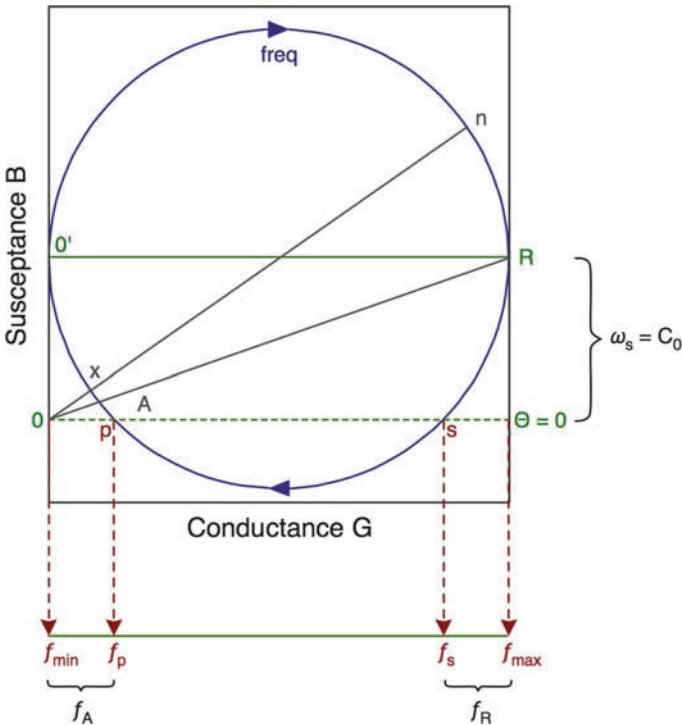

Fig. 4.15 Admittance diagram of quartz crystal microbalance (QCM) crystal and the relationship to characteristic frequencies (adapted from Yang and Thompson, 1993)
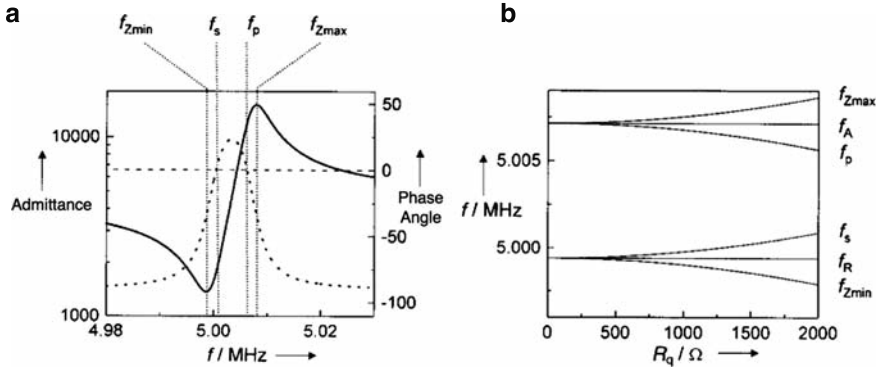
Fig. 4.16 (**a**) Bode plot; (**b**) effect of increasing dissipation (adapted from Janshoff et al., 2000, F8)

at which the conductance is at its minimum (f_{min}) and maximum (f_{max}), respectively. At the susceptance $B = \omega_s C_0$ the dotted line (zero phase angle) intersects the circle at parallel (f_p) and serial (f_s) frequencies. The crystal antiresonant (f_A) and resonant (f_R) frequencies are located between these values as $f_{min} > f_A > f_p$ and $f_s > f_R > f_{max}$. These relationships clearly demonstrate the fact that an overall resonant (or antiresonant) frequency of the crystal does not have an unambiguous diagnostic value, as the other frequencies do. In other words, mixed mass, elasticity, and dissipative effects contribute to the resonant frequency in a rather complex way.

Before we do the final extraction of the individual contributions we look at the impedance spectra, which are an equivalent format for performing the complex plane analysis (Fig. 4.16). The preference for performing the analysis in one or the other way is largely dictated by the instrument. In that case, the signal spectrum analyzer yields values of impedance Z and of the phase angle θ. Here we can again identify all the eigenfrequencies that we have seen in the Nyquist plot. Also, the fact that the definition of f_R and f_A is really "soft" is clearly evident in this representation. The meaning and the importance of the dotted line in Fig. 4.15 are also clearly seen. The line intersects the zero phase angle at two points and again defines the f_s and f_p, respectively.

The objective of the equivalent circuit modeling is to identify four unknowns, L, C, R, and C_0. Therefore, we need four equations. They are:

$$f_s = \chi \left(1 + \frac{C_0 R^2}{2L} \right) \tag{4.31a}$$

$$f_p = \chi \left(1 + \frac{C}{2C_0} - \frac{C_0 R^2}{2L} \right) \tag{4.31b}$$

$$f_{min} = \chi \left(1 - \frac{C_0 R^2}{2L} \right) \tag{4.31c}$$

$$f_{max} = \chi \left(1 + \frac{C}{2C_0} + \frac{C_0 R^2}{2L} \right) \tag{4.31d}$$

where

$$\chi = \frac{1}{2\pi}\sqrt{\frac{1}{LC}}. \qquad (4.31e)$$

The diagnostic power of the equivalent circuit analysis is seen in Fig. 4.16, in which the effect of increasing dissipation (resistance R) on the shape of the spectrum is clearly visible.

Even more revealing is the frequency dependence due to C, C_0, R, and L, as shown in Fig. 4.17.

In the sensing application, where the interaction of the selective layer with the analyte may lead not only to the mass uptake (i.e., Sauerbrey equation), but also to changes of its mechanical properties (i.e., elasticity and resistance), this form of analysis is the only way to extract meaningful information from QCM measurement. It is not simple, but it adds considerable information about the properties of the interface.
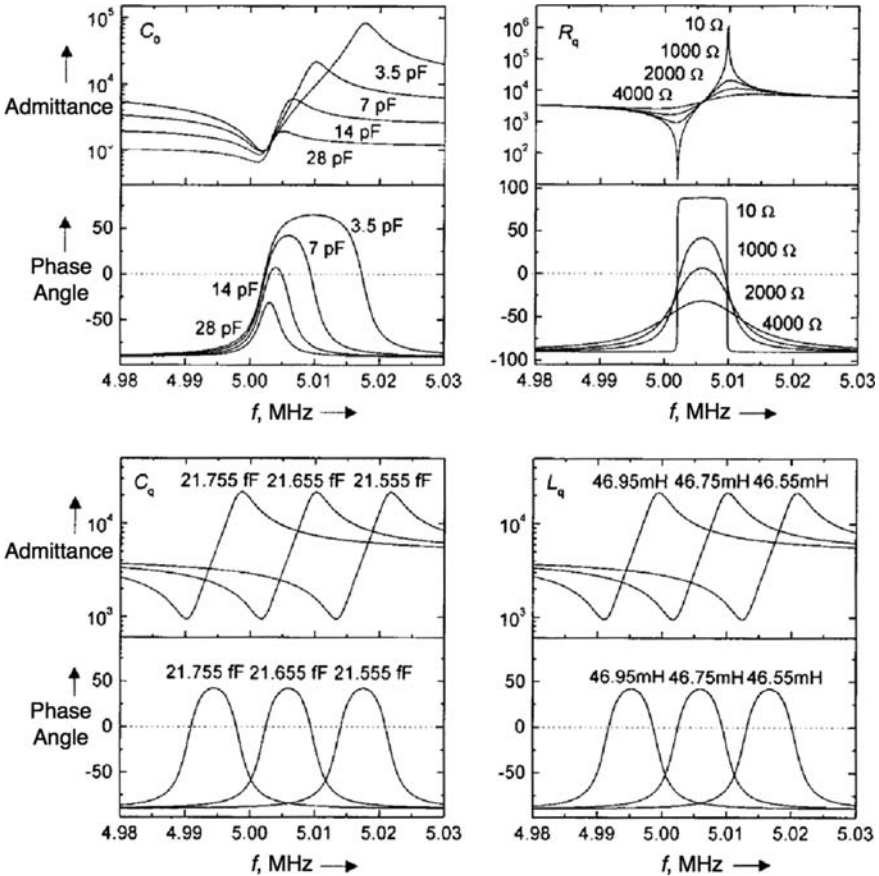
Fig. 4.17 Effect of frequency on C, C_0, R, and L (adapted from Janshoff et al., 2000, F7)

4.2.2 Surface Acoustic Wave Sensors

In 1885, Rayleigh predicted that surface acoustic waves could propagate along a
solid surface in contact with a medium of low density, for example, air (Ristic, 1983;
Grate et al., 1993). These waves, which are sometimes also called Rayleigh waves,
have considerable importance in areas as diverse as structural testing, telecommu-
nications, and signal processing. The exact description of the physical processes
involved in the generation, propagation, and detection of SAWs is complex and
beyond the scope of this book. Nevertheless, the potential of SAW devices as chemi-
cal sensors is considerable, despite the fact that the relationships between the output
signal of the SAW device with its associated electronics, and the added mass are
mostly empirical. In the following text, we concentrate on the physical processes
involved in the operation of these sensors.

   The simplest SAW sensor is a two-terminal transmission (delay) line in which the
acoustic (mechanical) wave is piezoelectrically launched in one oscillator, called the
transmitter. It travels along the surface of the substrate and is then transformed back
into an electrical signal by the reverse piezoelectric effect at the receiving oscillator
(Fig. 4.18).
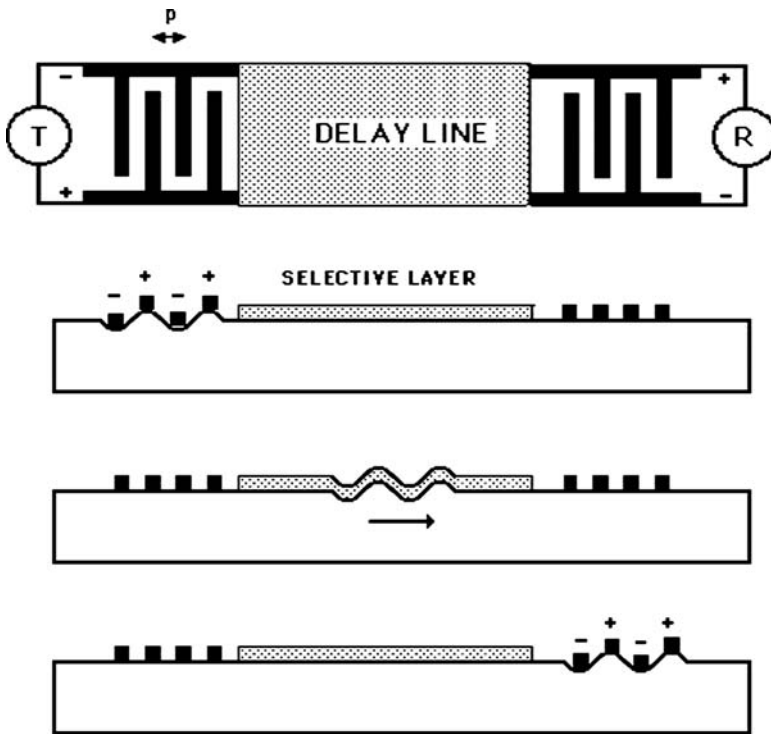
Fig. 4.18 Schematic of Surface Acoustic Wave (SAW) oscillator

From the point of view of chemical sensing, the analytical information is obtained from the interaction of the sample with the traveling wave in the region of the delay line. It is measured and related to concentration as attenuated amplitude ΔA_v, as phase difference (delay) $\Delta \phi$, or as frequency shift ΔF. The last transduction mode is the most common, because it is the most accurate and it allows direct comparison of sensitivities between different types of mass sensors. The transmitter and the receiver are interdigitated electrodes, deposited on top of the piezoelectric plate (Grate et al., 1993; Fig. 4.18). The frequency of the traveling wave is determined by the wave velocity v and by the center-to-center finger spacing p (see Fig. 4.18).

$$F = \frac{v}{p} \qquad (4.32)$$

The wave velocity depends on the plate material (typically quartz) and on the crystal cut. Typical frequencies are in the range of 30–300 MHz. Because we are dealing with the piezoelectric effect, the traveling wave has both mechanical (displacement) and electrical character. It is interesting to note that the wave traveling as a mechanical displacement has a velocity approximately five orders of magnitude lower than if it traveled the same distance as an electrical signal.

As shown in Fig. 4.19, in anisotropic medium, a surface acoustic wave represents a combined longitudinal (4.19a) and shear (4.19b) motion of the lattice in the y-(0)-z plane; this is the saggital plane. In anisotropic media, in certain multilayer structures and at some interfaces, the surface wave velocity exceeds the velocity of the shear wave and the energy continuously leaks from the surface to the bulk of the material. In such cases, we talk about pseudo- or "leaky" waves. Various energy-loss

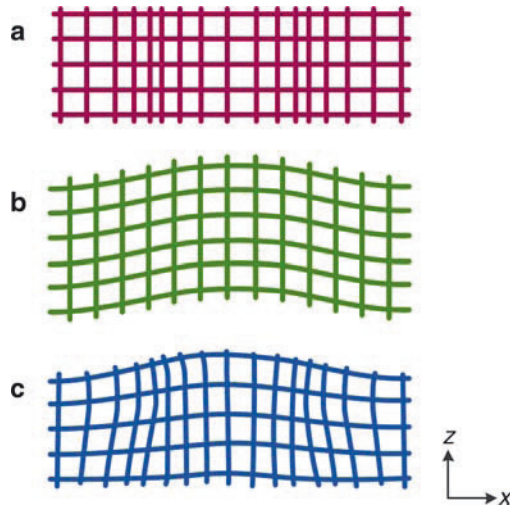

Fig. 4.19 Lattice distortion in Rayleigh wave defining the saggital plane: (**a**) longitudinal (bulk) wave; (**b**) y-polarized (shear bulk) wave; (**c**) y-0-z polarized Rayleigh waves (SAW waves) all propagating in the z-direction
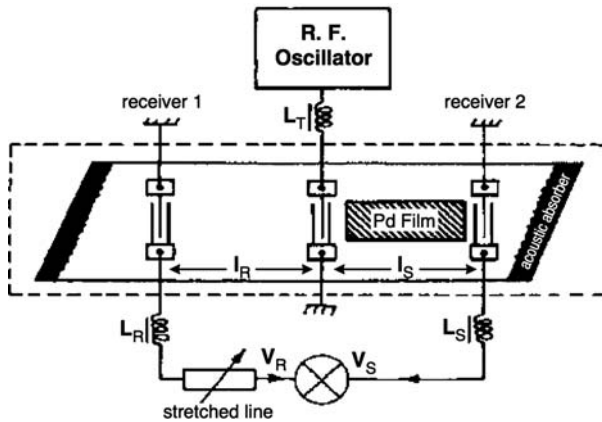
Fig. 4.20 Schematic of Surface Acoustic Wave (SAW) resonator for hydrogen sensing (adapted from D'Amico et al., 1982/83)

mechanisms operate on the propagating SAW. Scattering loss due to the granularity of the deposited layer, thermoelastic loss due to the nonadiabatic behavior of the acoustic conductor, viscous loss caused by the dissipation of energy in the direction normal to the liquid–solid interface, and hysteresis absorption due to irreversible coupling of the SAW energy to the adsorbate.

Most SAW sensors employ a resonator arrangement: a sensor delay line and the reference line are combined on one substrate with one generator in the center and two receivers at the ends (D'Amico et al., 1982/83; Fig. 4.20) terminated by two reflectors.

The usual substrate materials from which SAW devices are made are ST-cut (stress- and temperature-compensated) quartz or $LiNbO_3$, which has a high piezoelectric coefficient. The area between the transmitter and the receiver, the delay line, does not have to be made from the same material; as a matter of fact, it does not have to be even piezoelectric. In that case, acoustic couplings are used to transfer the energy from the piezoelectric transmitter/receiver to and from the delay line. Combination of all these factors results in a perplexing number of possibilities for the design of practical SAW sensors and for the ways in which the output signal can be obtained. Integrated circuit techniques can be used for preparation of these devices. It is also possible to represent the SAW sensor by an equivalent electrical circuit and to perform some optimization prior to fabrication in order to minimize the energy losses. The scaling relationships that apply to the SAW sensors are summarized in Table 4.1.

The phase shifts in the sensing ϕ_s and in the reference ϕ_r lines are

$$\phi_r = \frac{2\pi F l_r}{v_r} + \phi_E \tag{4.33}$$

and

$$\phi_s = \frac{2\pi F l_s}{v_s} \tag{4.34}$$

Here, ϕ_E is the phase shift introduced by the compensating (stretched) line, l_r and l_s are the length of the reference and of the sensing channel, respectively, and v_r and v_s are the acoustic velocities in the two channels. The output of the sensor (V_0) is then

$$V_0 = V_M \sin(\phi_r - \phi_s) \tag{4.35}$$

where V_M is the maximum output, occurring at $(\phi_r - \phi_s) = k\pi/2$ (for $k = 1, 3, \ldots$, $2n - 1$). The stretched line is used to set the phase shift difference for zero concentration. Equations. (4.33) through (4.35) yield

$$V_0 = V_M \sin\left[2\pi F \left(l_r/v_r - l_s/v_s\right) + \phi_E\right] \tag{4.36}$$

The addition of mass due to the interaction with the sample causes the change of the acoustic velocity v_s in the sensing line. Therefore, the sensitivity S of the device is obtained by differentiating (4.36) with respect to the concentration of gas C_{gas}.

$$S = \frac{dV_0}{dC_{gas}} = -\frac{2\pi V_M F l_s}{v_s^2} \left(\frac{dv_s}{dC_{gas}}\right) \cos\left[2\pi F \left(l_r/v_r - l_s/v_s\right) + \phi_E\right] \tag{4.37}$$

By using the stretched line, the phase-shift (ϕ_E) can be adjusted in such a way that the argument in the cosine is zero, in which case (4.37) reduces to

$$S = \frac{dV_0}{dC_{gas}} = -\frac{2\pi V_M F l_s}{v_s^2} \left(\frac{dv_s}{dC_{gas}}\right) \tag{4.38}$$

Temperature and pressure also affect the SAW velocities. The pressure sensitivity (S_P) and temperature sensitivity (S_T), respectively, are

$$S_P = \frac{dV_0}{dP} = 2\pi V_M F \left(\frac{l_s dv_s}{v_s^2 dP} - \frac{l_r dv_r}{v_r^2 dP}\right) \tag{4.39}$$

and

$$S_T = \frac{dV_0}{dT} = 2\pi V_M F \left(\frac{l_s dv_s}{v_s^2 dT} - \frac{l_r dv_r}{v_r^2 dT}\right) \tag{4.40}$$

Thus, a proper choice of l_s and l_r can minimize the pressure and temperature dependence of these devices.

An example of a hydrogen SAW sensor is shown in Fig. 4.20, where the area of L_s is coated with 3,000 Å of Pd. The substrate is LiNbO$_3$. There are five finger pairs and the device is operated at $F_0 = 75\,\text{MHz}$. The width of the acoustic beam is 100 acoustic wavelengths (4.1). The sensor responds to change of hydrogen concentration in the range 50–10,000 ppm.

Various thin selective layers, as discussed in Chapter 1, can be used to provide chemical selectivity. As with piezoelectric crystals, the acoustic properties of these films affect the performance of the SAW sensor in different ways: by the change of

mass due to the adsorbed/absorbed gas or by the change of the mechanical properties of the film. An electrostatic mode of coupling between the chemically selective coating and the piezoelectric delay line has been also observed (Ricco et al., 1985). This electroacoustic effect can happen when the film is a semiconductor with a low density of charge carriers. As the piezoelectric wave propagates through the delay line, it induces a traveling electrical wave in the adjacent film whose velocity depends on the conductivity of the film. That, in turn, can be modulated by the interaction of the film with an electron-donating or electron-accepting gas.

A comparative study of the readout options for the SAW sensor with additional film has shown that for a single SAW sensor the highest signal-to-noise ratio is obtained from the amplitude measurement (Wohltjen and Dessy, 1979). Voltage output related to the phase-shift as discussed above works well for dual delay lines. There are also inherent advantages in measurement of the change of the resonant frequency. The frequency shift due to deposited film of low elastic shear modulus μ is

$$\Delta F = (k_1 + k_2)F_0^2 t_f \rho - k_2 F_0^2 t_r \left[\frac{4\mu(\chi_L + \mu)}{V_r^2(\chi_L + 2\mu)} \right] \qquad (4.41)$$

where k_1 and k_2 are material constants of the SAW substrate, t_f is the film thickness and ρ is its density, χ_Λ is the Lamé constant, and μ is the elastic modulus of the film. Thus, the film mass/unit area is $t_f \rho$. If the chemical interaction does not change the mechanical properties of the film, the second term in (4.41) can be neglected and the frequency shift ΔF is due exclusively to the added mass (Wohltjen et al., 1979). The absolute and relative shifts are described, respectively, in (4.42a) and (4.42b).

$$\Delta F = (k_1 + k_2)F_0^2 t_f \rho \qquad (4.42a)$$

$$\frac{\Delta F}{F} = (k_1 + k_2)F \Delta M_f \qquad (4.42b)$$

Thus, the addition of a 1 μm thick polymer film coated onto a quartz SAW sensor operating at 31 MHz will cause a $\Delta F = -130$ kHz shift.

There are many similarities and some differences between QCM and SAW devices that should be noted. First of all, both types of sensors transduce the changes in mass to electrical signal, with SAW's being approximately one order of magnitude more sensitive (Table 4.2). It is their primary function as chemical sensors. The same selective layers and methods of deposition can be used with both apparatuses. However, here comes the first major difference. Because in QCM the selective layer lies on a metal electrode, it can be deposited and/or modified electrochemically. This is particularly convenient if, for example, conducting polymers are used as selective layers. On the other hand, such a layer can also generate the electroacoustic effect in the delay line of the SAW which, in principle, can be also utilized for chemical sensing. Both types of devices couple the mechanical motion to the surrounding medium. This coupling is particularly pronounced in operations in liquids, where it can yield additional information about the dynamics of the sample–layer interface or it can be a source of major artifacts. Equivalent electrical circuit modeling can help and it has been particularly well studied and developed for QCMs.

The packaging (i.e., electrical insulation for operation in electrolytes) is more difficult with SAWs due to their rectangular geometry. SAWs are easier to fabricate with lithographic microfabrication techniques and therefore are more suitable for use in an array (Ricco et al., 1998). The choice of electrode materials is critical for QCM, where acoustic impedance mismatch can result in substantial lowering of the Q factor of the device. On the other hand, it does not play any role in the SAW devices. The energy losses to the condensed medium are higher in SAWs and this fact makes them even less suitable for operation in liquids. Nevertheless, SAW biosensors have been reported (Marx, 2003).

4.2.3 Plate Mode Oscillators

In the devices discussed above, the acoustic wave propagates in a slab of piezo-electric material whose thickness is infinitely larger than the wavelength of the propagating wave (Table 4.2). When the thickness of the plate is reduced so that it becomes comparable to λ, the whole plate becomes involved in the periodic motion and a symmetric and antisymmetric Lamb wave is created (White et al., 1987; Fig. 4.21).

This happens in Acoustic plate-mode (Apm) oscillators, which have a thickness of only a few microns. The eigenfrequency of the Apm oscillators is given by the interdigitated electrode spacing p and by the plate thickness t.

$$F_0 = \frac{v}{p} \left[1 + \left(\frac{np}{2t} \right)^2 \right]^{1/2} \tag{4.43}$$

There are certain aspects of performance that make the Apm oscillators potentially attractive as chemical sensors. First of all, the fact that both surfaces contribute to the signal means that the sensitivity is higher than for the corresponding SAW device. The most important advantage follows from the fact that velocity of the lowest order of the antisymmetric mode is much slower than the compressional velocity of sound in most liquids ($900–1,500 \, \text{m s}^{-1}$), which means that the energy
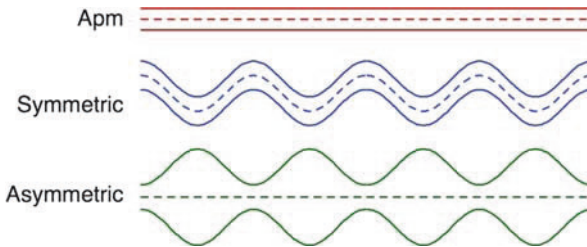

Fig. 4.21 Bulk Lamb waves in Acoustic plate-mode (Apm) and Flextural Plate Wave (FPW) oscillators

losses in the direction normal to the surface of the plate are minimized. This is important for potential applications to mass sensing in liquids.

Flextural Plate Wave (FWP) resonators belong to this family of devices. They are fabricated from a few microns thick rectangular membranes of $Si_3 N_4$ or ZnO that can be (but do not have to be) piezoelectric. The Lamb waves are excited piezoelectrically (e.g., ZnO), either electrostatically or electromagnetically. The fundamental frequency is given by

$$F_0 = \frac{\sqrt{t_B}}{\lambda M}$$

(4.44)

where t_B is the bending thickness of the membrane and M is its mass per unit area. There are some practical problems with the use of these devices, the fragility of the thin plate and the sensitivity to the external pressure being the most serious. The Lamb wave also has a higher frequency dispersion than the corresponding surface acoustic wave.

4.3 Nonpiezoelectric Mass Sensors

The latest additions to the family of mass sensors are *cantilevers*. As has been the case with other types of mass sensors, they were originally developed for an entirely different application: as the tips for scanning probe microscopies, namely the atomic force microscopy (Binning et al., 1986). Only as an "afterthought" was it discovered (Gimzewski et al., 1993) that both static and dynamic properties of these tips change with the chemical environment in which they operate. Unlike the rest of mass sensors, cantilevers are not built from nor do they depend on piezoelectric materials. They are microfabricated objects, usually from a highly inert material, such as silicon nitride. The shape and dimensions of cantilevers are design parameters and vary from application to application. The motion of the cantilever is detected optically, using a Position-Sensitive Detector (PSD), as shown in Fig. 4.22.

There are two modes in which cantilevers can function as chemical sensors (Fig. 4.22). These are the static, or *bending* mode and the dynamic, or *resonant* mode.

In the bending mode, the adsorption of material at one surface of the cantilever causes the change of the surface energy (stress) and bending, according to Stoney's law.

$$\sigma = \frac{Et^2}{6R_B(1 - \xi_P)}$$

(4.45)

where σ is stress, the Poisson ratio $\xi_P = \Delta A / \Delta l$, t is thickness, E is the Young's modulus, and R_B is the bending radius. The deformation parameters for silicon are $E = 1.7 \times 10^{11} \text{N m}^{-2}$, and $\xi = 0.25$. The analytical information is obtained from the calibration curve. There is no gravitational or inertial mass involved in this measurement. The transduction principle is based on asymmetrical change of the surface energy and the cantilever functions in a similar fashion as a bimetallic strip. It is an
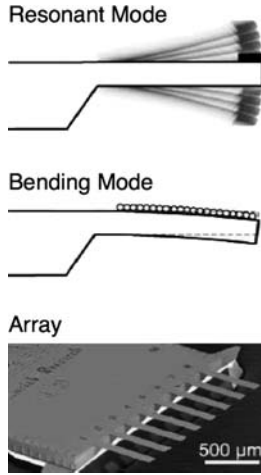
Fig. 4.22 Schematic of cantilever operating in (**a**) resonant and (**b**) static bending mode. (**c**) Photograph of microfabricated cantilever array (adapted from Battiston et al., 2001, p. 122)

interesting, but rather unusual sensing principle that is not discussed further. It is mentioned in the context of mass sensors only because the cantilever and the attendant instrumentation are identical when they are used in the resonant mode, which depends on the inertial mass and is therefore a true mass-sensing device.

4.3.1 Resonant Cantilevers

The material properties and geometry of the cantilever determine its spring constant and ultimately its resonant frequency. Because they are long and thin objects of width w for which $l >> t$, the usual approximate expression for the spring constant K is

$$K = \frac{Et^3w}{4l^3} \tag{4.46}$$

The resonant frequency is then

$$F_0 = \frac{tl^2}{2\pi(E/\rho)^{1/2}} \tag{4.47}$$

The deposition of mass ΔM results in decrease of the resonant frequency according to

$$\Delta M_f = \frac{K(F^{-2} - F_0^{-2})}{4n_c\pi} \tag{4.48}$$

The geometrically dependent correction factor n_c has value 0.24 for the most common rectangular cantilevers. Equation (4.48) is valid only if the spring constant K

does not change during the experiment. It is a condition that may not be satisfied if, for example, the elasticity of the selective layer changes upon exposure to the sample. In that respect, it is an analogous problem that was encountered with practical QCMs and SAWs.

As can be seen from (4.46) through (4.48), the mass sensitivity depends on the geometry of the beam and can be optimized for most applications. For a typical dimension of a cantilever fabricated from silicon, $l = 500\,\mu m$, $w = 100\,\mu m$, and $t = 8.6\,\mu m$, the resonance frequency is approximately $50\,kHz$. Here comes another advantage similar to the one that FPW and Apm oscillators have. Their frequency is well below the velocity of acoustic compressional waves in most liquids. This means that the problem of energy dissipation and coupling to liquid media is greatly diminished. In fact, it is possible to use the resonant frequency of the cantilever for biosensing and for the determination of the Reynolds number (Chon et al., 2000).

$$R_n = \frac{\rho \pi F w^2}{2\eta} \tag{4.49}$$

Because the vibration is spontaneous (i.e., thermally actuated) and the readout is optical (Fig. 4.23), there are no electrical connections required for their operation even in conducting electrolyte solutions. This greatly simplifies the packaging and encapsulation for biosensing, for example.

Inherent in the oscillation of any object is the existence of different modes of oscillations and of higher harmonics. In addition to the lateral modes there is also the torsional mode. All modes can be detected in the x–y-directions by the PSD (Fig. 4.22). In the case of resonant cantilevers, the different modes have different mass sensitivity (Kim et al., 2001). They are well separated on the amplitude spectrum (Fig. 4.24), and show different sensitivities to absorption of vapors into the selective coating applied to the cantilever (Fig. 4.25).
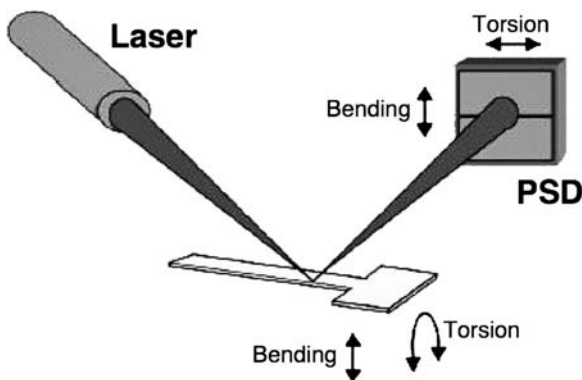
Fig. 4.23 Optical readout of motion of cantilever in x–y direction by 2D position sensitive detector (adapted from Kim et al., 2001, p. 12)
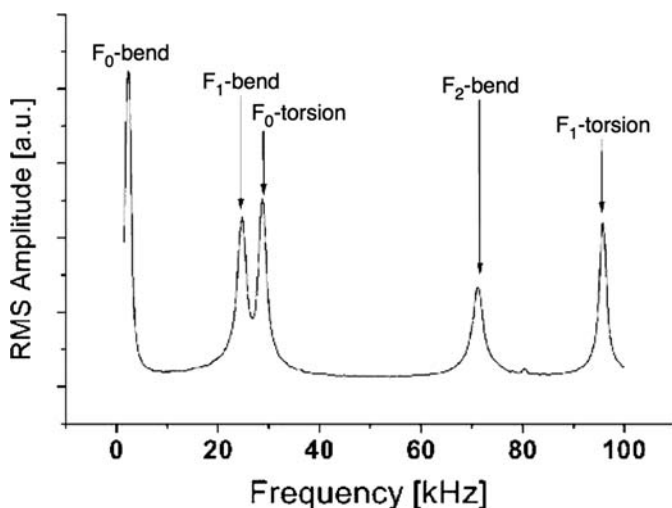
Fig. 4.24 Root-mean-square (RMS) amplitude spectrum of cantilever exhibiting lateral and torsional vibration modes and higher harmonics (adapted from Kim et al., 2001)
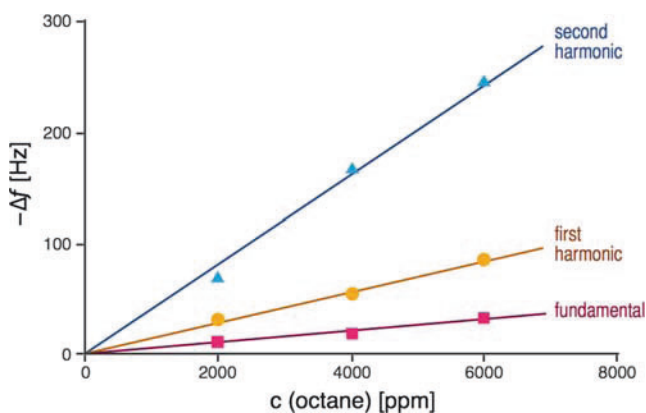
Fig. 4.25 Illustration of mass sensitivities to sorption of octane by poly(dimethylsiloxane) coating, of different vibrational modes (adapted from Kim et al., 2001)

Because of this great flexibility in the design, fabrication, and readout, cantilever arrays for gaseous (Kim et al., 2001) and biosensing aqueous applications (Arntz et al., 2003) have been realized. Needless to say, the issues of selectivity, dynamic range, response time, and so on depend ultimately on the interactions of the analyte with the selective layer. What cantilevers can offer is a low instrumental detection limit and the possibility of avoiding experimental artifacts due to the effects not related to the chemical interactions. Because they are relatively new, it is too early to estimate their usefulness. They will have to stand the test of time.

Food for Thought #4

Electrochemical Calibration of QCM

Electrochemical calibration of QCM using the electroplated metals according to the Faraday's law has been reported.

1. Can such an approach be used for verification of the Sauerbrey equation? Discuss the origin of possible experimental artifacts.

Gravitational Effects

2. Imagine that NASA has given you a ton of money to develop a mass sensor for the International Space Station. It will be used for monitoring the air quality inside the station. Which type of mass sensor could you NOT use: QCM, SAW, cantilever (in bending), or cantilever (resonant mode)?

Enzymatic Mass Sensors

3. It is generally true that catalytic selectivity is not a good basis for mass sensors. Propose a scheme in which a highly selective interaction between the enzyme and the substrate could be used to detect interaction as a change in mass.

Determination of Hydration Numbers

4. Let us assume that some interactions involving biomolecules will not take place in a completely dry state. Could you design an experiment that would allow you to determine the minimum number of water molecules required for such interaction to take place?

Impedance Matching

5. The requirement of impedance matching between quartz and the metal electrodes is very important for the operation of the QCM. Can you think of an arrangement that would be completely independent of the type of metal used to drive the crystal?
6. For what type of sensing applications could you use it, liquids or gas?
7. Make a sketch of the proposed arrangement and include realistic operational parameters (i.e., voltages, dimensions, mounting arrangements).

Symbols

ΔA_v	Attenuated amplitude
B	Susceptance
C_{gas}	Concentration of a gas
C	Capacitance
C_q	Crystal constant
D_{crit}	Optimum diameter
E	Young's modulus
E_c	Collection efficiency
F	Linear frequency; linear frequency
F_f	Resonant frequency of the crystal with material deposited on it
G	Conductance
H	Transfer function
I	(or i) Current
j	Standard notation for an imaginary number
k	Material constant
K	Spring constant
l	Length
L	Inductance
M	Mass
n_c	Geometrically dependent correction factor
p	Center-to-center finger spacing
P	Gas pressure
R	Resistance
R_n	Reynolds number; a fluid mechanics term. It is a dimensionless ratio of inertial and viscous properties of the fluid. It is a figure of merit for a given fluid.
S	Sensitivity
t	Thickness
v	Shear velocity
V	Voltage
w	Width
Y	Admittance
Z	Acoustic impedances
δ	Damping distance
κ_L	Kinematic viscosity
π	Mathematical constant *pi* (3.1415...)
λ	Wavelength
μ	Shear modulus
ρ	Density
η	Viscosity
θ	Phase angle
ϕ	Phase shift
σ	Stress

ω	Angular frequency
ξ_P	Poisson ratio
ζ	Parameter defining the acoustic matching of a crystal and film material
χ_Λ	Lame constant

References

Arntz, Y., Seelig, J.D., Lang, H.P., Zhang, J., Hunziker, P., Ramseyer, J.P., Meyer, E., Hegner, M., and Gerber, C. (2003) *Nanotechnology* 14, 86.

Battiston, F.M., Ramseyer, J.-P., Lang, H.P., Baller, M.K., Gerber, C., Gimzewski, J.K., Meyer, E., and Güntherodt, H.-J. (2001) *Sens. Actuators B* 77, 122.

Binning, G., Quate, C.F., and Gerber, C. (1986) *Phys. Rev. Lett*. 56, 930.

Bizet, K., Gabrielli, C., Perrot, H., and Therasse, J. (1998) *Biosens. Bioelectron*. 13, 259–269.

Bruckenstein, S., Michalski, M., Fensore, A., Zhufen, K., and Hillman, A.R. (1994) *Anal. Chem*. 66, 1847.

Chon, J.W.M., Mulvaney, P., and Sader, J.E. (2000) *J. Appl. Phys*. 87, 3978.

D'Amico, A., Palma A., and Verona, E. (1982/83) *Sens. Actuators* 3, 31.

Dunham, G.C., Benson, N.H., Petelenz, D., and Janata, J. (1995) Dual quartz crystal microbalance. *Anal. Chem*. 67, 267–272.

Gimzewski, J.K., Gerber, C., Meyer, E., and Schlittler, R.R. (1993) *Chem. Phys. Lett*. 217, 589.

Grate, J.W., Martin, S.J., and White, R.M. (1993) *Anal. Chem*. 65 940A; 987A.

Hillier, A.C. and Ward, M.D. (1992) *Anal. Chem*. 64, 2539.

Ho, M.H. (1984) Application of quartz crystal microbalances in aerosol mass measurement. In: C. Lu and A.W. Czaderna (Eds.) *Methods and Phenomena, Vol. 7*. Elsevier.

Janshoff, A., Galla, H.-J., and Steinem, C. (2000) *Angew. Chem. Int. Ed*. 39, 4004–4032.

Kanazawa, K.K. and Gordon, J.G. (1985) *Anal. Chem*. 57, 1770.

Kim, B.H., Prins, F.E., Kern, D.P., Raible, S., and Weimar, U. (2001) *Sens. Actuators B* 78, 12.

King, Jr., W.H. (1964) *Anal. Chem*. 36, 1735.

Lin, Z. and Ward, M.D. (1995) *Anal. Chem*. 67, 685.

Lu, C. and Czaderna, A.W. (Eds.) (1984) *Methods and Phenomena*, *Vol. 7*. Elsevier.

Martin, B.A. and Hager, H.E. (1989a) *J. Appl. Phys*. 65, 2627.

Martin, B.A. and Hager, H.E. (1989b) *J. Appl. Phys*. 65, 2630.

Mason, W.P. (1965) *Physical Acoustics*, *Vol. 2A*. Academic Press.

Marx, K.A. (2003) *Biomacromolecules* 4, 1099–1120.

Ricco, A.J., Martin, S.J., and Zipperian, T.E. (1985) *Sens. Actuators* 8, 319.

Ricco, A.J., Crooks, R.M., and Osbourn, G.C. (1998) *Acct. Chem. Res*. 31, 289.

Ristic, V.M. (1983) *Principles of Acoustic Devices*. Wiley.

Schneider, T.W. and Martin, S.J. (1995) *Anal. Chem*. 67, 3324.

Theisen, L.A. Martin S.J., and Hillman A.R. (2004) *Anal. Chem*. 76, 796–804.

White, R.M., Wicher, P.J., Wenzel, S.W., and Zellers, E.T. (1987) *IEEE Trans. Ultrason. Dev. Ferroel. Freq. Contr*. UFFC-34, 162.

Wohltjen, H. and Dessy, R. (1979) *Anal. Chem*. 51, 1465.

Yang, M. and Thompson, M. (1993) *Anal. Chem*. 65, 1158.

Chapter 5
Electrochemical Sensors

5.1 Introduction

Electrochemical sensors are the largest and the oldest group of chemical sensors. Some have reached commercial maturity, and others are still in various stages of development. Thorough coverage of this important sensor group requires a four-chapter block of this text. In Chapters 5 through 8, we examine sensors as diverse as enzyme electrodes, high-temperature oxide sensors, fuel cells, and surface conductivity sensors. Each is discussed in the broadest meaning of the word "electrochemistry," that is, the interaction between electricity and chemistry.

These chapters divide the discussion of electrochemical sensors by the mode of measurement. This chapter is an introduction to the general parameters and characteristics of electrochemical sensors. Chapter 6 focuses on potentiometric sensors, which measure voltage. Chapter 7 describes amperometric sensors, which measure current. Chapter 8 examines conductometric sensors, which measure conductivity.

Despite the pervasive use of electrochemical sensors and the fundamental importance of electrochemistry as a division of physical and analytical chemistry, this field of study has not traditionally been a favorite of students. One reason for this could be the fact that most electrochemical and electroanalytical textbooks introduce electrochemistry by explaining first the thermodynamics of the electrochemical cell. That approach is bound to discourage all but the brave few.

We take a different approach and first examine what happens inside the electrochemical cell when current passes through it. It is done in three simple "Thought Experiments." Because of the large number of possible variations in the arrangement of an electrochemical experiment, it is not possible to construct a generalized cell with corresponding general current–voltage characteristics that would be representative of any and all configurations. The experimental conditions play a decisive role in the signal output. Our Thought Experiments are selected to explain the basic electrochemical relationships, and elucidate expected results. Before we proceed with the Thought Experiments, however, there are a few general parameters of all electrochemical cells that must first be understood.

5.2 General Rules of Electrochemistry

Electrochemistry implies the transfer of charge from the electrode to another phase, which can be a solid or a liquid. During this process, chemical changes take place at the electrodes and the charge is conducted through the bulk of the sample. Both the electrode reactions and/or the charge transport can be chemically modulated, and can serve as the basis of the sensing process. Although many different cell arrangements can be devised, there are three rules that apply to all electrochemical sensors.

Rule #1. The first rule is the *requirement of the closed electrical circuit.* This means that at least two electrodes must be present in the electrochemical cell. From a purely electrical point of view, it means that we have a sensor electrode (the working electrode) and a signal return electrode (often called the auxiliary electrode). This requirement does not necessarily mean that a DC electrical current will flow in a closed circuit. Obviously, if we consider an ideal capacitor **C** in series with a resistor **R** (Appendix C), a DC voltage will appear across the capacitor, but only as a transient; DC current will not flow through it. On the other hand, if an AC voltage is applied to the cell, a continuous displacement charging current will flow.

An "open" electrical circuit (such as an "electrochemical cell" consisting of only one electrode!) simply means that a small and undefined capacitor (the "missing electrode") has been placed in series in the circuit. The result is predictable: it will readily respond to high-frequency electrical fluctuations (e.g., noise), but no information that depends on DC behavior (i.e., on the composition of the sample) can be obtained.

This is not necessarily a problem. For example, an antenna transmitting at high frequency can be regarded as one electrode, whereas the receiver (the "other electrode") does not have to be even physically defined. In the extreme case, it can be the Earth. If a dielectric object is placed in the proximity of the transmitter, the amount of reflected radio-frequency energy depends on the average dielectric constant of the object. This is the basis of one of the most accurate, continuous humidity sensors because the dielectric constant of the sample is a strong function of the content of water. Admittedly, such an "electrochemical sensor" is at the boundary of the scope of this book, but it clearly illustrates the point that the frequency of the applied signal is an important parameter in considering the continuity of the electrical circuit.

Rule #2 The second rule is the *condition of electroneutrality*. It means that in an electrochemical cell, the sum of positive charges must equal the sum of negative charges. Thus, separation of positive and negative charges occurs at every interface, but their sum is always zero.

Rule #3 This is not so much a rule as it is an important general point regarding the nature of the interfacial reaction in all electrochemical sensors. Charge transport within the transducer part of the sensor, and/or inside the supporting instrumentation, is *electronic*. On the other hand, the charge transport in the sample can be *electronic, ionic,* or *mixed (electronic/ionic)*. In the latter two cases, an electron

transfer at the electrode/sample interface and its mechanism becomes one of the most critical aspects of sensor performance. If a nonzero current flows through this electronic/ionic interface, a chemical transformation – electrolysis – takes place, according to Faraday's law (5.1). The chemical change is caused by the passage of charge Q through the electrochemical interface, at which the electronic conduction in the metal changes into ionic conduction in the sample.

$$m = \frac{Q}{nF} \tag{5.1}$$

In (5.1), the symbol F is the Faraday constant (96,493 coulombs) and m is the number of moles converted in a reaction typified as (5.2).

$$O + ne \underset{k_a}{\overset{k_c}{\Leftrightarrow}} R \tag{5.2}$$

Here **O** represents the oxidized species, **R** the reduced species, and n is the number of electrons. Species **O** and **R** differ only by n electrons; they are called a redox couple. The rate constants k_c and k_a describe the dynamic nature of the reduction (at the cathode) and oxidation process (at the anode). When the passage of current through the interface results in measurable changes of bulk concentration, it is convenient to write Faraday's law in terms of molar concentration C and volume V_{cell} of the cell.

$$C = \frac{Q}{nFV_{cell}} \tag{5.3}$$

A schematic diagram of a basic electrochemical cell is shown in the insert of Fig. 5.1. It contains a working electrode **W**, auxiliary electrode **AUX**, signal source **S** (current or voltage), and meter **M** for measurement of voltage or current. The ionic medium in which the electrodes are immersed is an electrolyte, which can be either liquid or solid.

   The overall current–voltage relationship is complex and very different for different experimental conditions. In Thought Experiment I, we begin to examine its characteristics, and continue to look deeper in Experiments II and III. These exercises help to define the general parameters discussed thus far in specific operational scenarios. For all three Thought Experiments, consult Fig. 5.1.

5.2.1 Thought Experiment I (Curve A)

Let us apply potential E from signal source **S**, between working **W** and auxiliary **AUX** platinum electrodes, starting from zero volts and moving in the negative direction. Platinum is selected as the electrode material, because it is electrochemically inert (i.e., nonreactive). The electrolyte (i.e., ionically conducting phase) is an aqueous solution of H_2SO_4 and is stirred. The polarity refers to the direction of electron transfer through the cell. Said another way, at which electrode are electrons
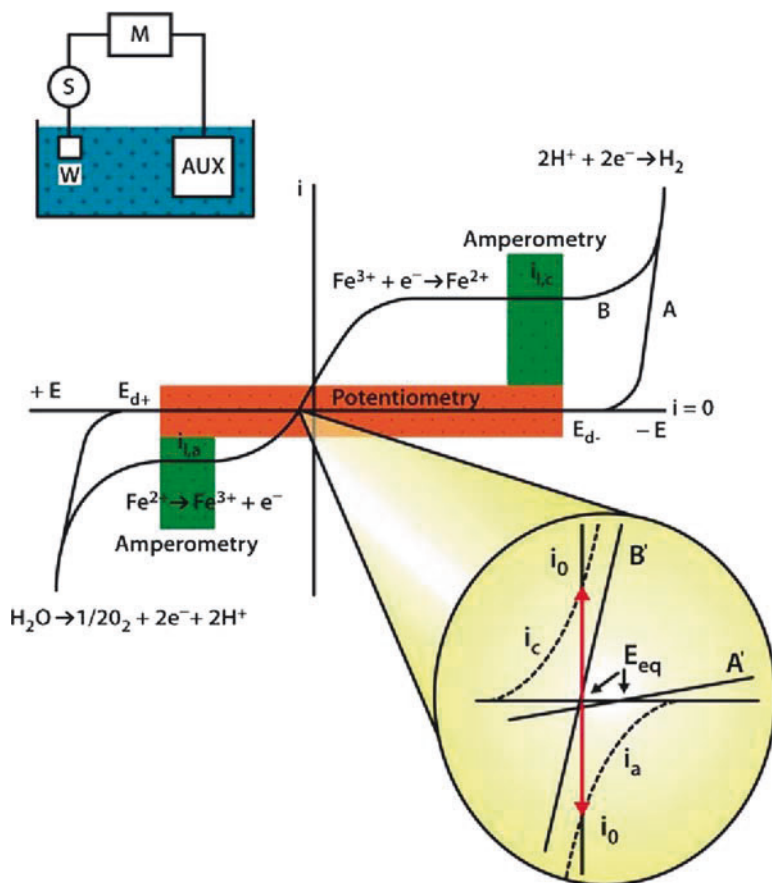
Fig. 5.1 Current–voltage curves obtained in a generalized electrochemical experiment

consumed by the ionic species (the cathode) and at which electrode are electrons liberated from the ionic species (the anode)? This is determined by which electrode has a potential applied to it. At this point in the Thought Experiment, we designate **W** as the cathode. Also note that **W** has a much smaller surface area than **AUX**. The reasons for this choice is explained later.

At first, no current flows through the cell until a decomposition potential E_{d-} is reached. At that point, the current begins to flow (Curve A). We also observe that gas bubbles are formed at the working electrode, and that current fluctuates somewhat randomly. Two chemical processes are taking place at the electrodes. At the **W** electrode (which is now the cathode), electrons are transferred from the electrode to the hydrogen ions, H^+. Thus, the reduction takes place at the cathode, according to the following electrochemical reaction.

$$2H^+ + 2e^- \Rightarrow H_2 \tag{5.4}$$

Because the electrical circuit is closed, the opposite (oxidation) electrochemical reaction takes place at the larger **AUX** electrode, where the electrons leave the solution. So, at the anode, we have

$$H_2O \Rightarrow \frac{1}{2}O_2 + 2e^- + 2H^+ \tag{5.5}$$

Thus, for the transfer of two electrons through this electrochemical circuit, one molecule of water is decomposed (electrolyzed) to half a mole of oxygen (at the anode) and one mole of hydrogen (at the cathode).

In order to record the positive half of Curve A on the same current scale, we have to change the polarity of the applied voltage to **W**, making it the anode. That is, we start again from zero volts and slowly increase the voltage to $+2\,V$. Then, of course, the electrons are withdrawn from the solution at the working electrode, whereas at E_{d+} the oxidation according to (5.5) takes place. The voltage between the two decomposition potentials represents the electrochemical window $<E_{d+} : E_{d-}>$ in which the electrochemical measurements can be made. Note that voltage applied to the cell is always the difference of two potentials at individual electrodes.

This simple experiment contains a wealth of electrochemical information which we now summarize in three interim conclusions.

1. *The current–voltage relationship in the electrochemical cell is nonlinear and is not governed by Ohm's law $(I = E/R)$. There is a region of voltage where no current flows. It means that there must be at least one capacitor* $\mathbf{C_{dl}}$ *in series, in the cell. Actually, there are two and they are located at the interfaces of the* **W** *and* **AUX** *electrodes with the electrolyte solution. They store energy according to the applied voltage across them. The cell, therefore, represents a capacitive divider. When the voltage exceeds the dielectric breakdown potential of these capacitors, either at the positive* (E_{d+}) *or at the negative* (E_{d-}) *threshold, current begins to flow. At that point, the capacitor is shorted out by a resistor* $\mathbf{R_{ct}}$ *in parallel. There is also a resistor* $\mathbf{R_s}$ *in series, which represents the resistance of the electrolyte solution.*
2. *Random fluctuations of the current correlate with the evolution of the bubbles at the working electrode (see FFT).*
3. *Both R_{ct} and C_{dl} are voltage-dependent, but R_S is not.* Explanation of the location, and properties of $\mathbf{R_{ct}}$, $\mathbf{R_s}$, and $\mathbf{C_{dl}}$ is what electrochemistry is about. We explain these nonlinearity-causing resistors and capacitors by equivalent electrical circuit analysis as we go forward.

5.2.2 Thought Experiment II (Curve B)

Let us now add some soluble redox couple, such as a mixture of Fe^{2+} and Fe^{3+} in approximately $10^{-3}\,M$ concentrations and record the current–voltage curve again while stirring (Curve B). This redox couple is defined by the reaction shown in (5.6).

$$\mathrm{Fe^{3+}} + e \underset{k_\mathrm{a}}{\overset{k_\mathrm{c}}{\Leftrightarrow}} \mathrm{Fe^{2+}} \tag{5.6}$$

The reason why Curve B looks different from Curve A is that the added redox couple allows the electrons to jump across the interface (capacitor) at much lower energy (voltage) than before. As a result, it shunts (depolarizes) the capacitor. Thus, the energy stored in the capacitor has been used to oxidize or reduce the added depolarizer in the solution. We can make four observations from Thought Experiment II.

1. *Current–voltage Curve B is even more nonlinear than Curve A.* It intercepts the zero current line, or open cell potential line, at a single point, at which $i = 0$. That point is the equilibrium potential E_eq of the above redox couple.
2. *Upon making the potential of the working electrode slightly more negative with respect to E_eq, the current increases due to the reduction of Fe^{3+} ions* (5.6). Similarly, upon increase of the potential in the positive direction from E_eq, ferrous ions ($\mathrm{Fe^{2+}}$) are oxidized, according to (5.6).
3. *A further increase of the applied voltage in either direction causes the current to reach the limiting current plateau.* Within this window, reduction and oxidation of ferric and ferrous ions, respectively, are taking place at a constant rate.
4. *Upon further increase of voltage, the current reaches $+E_\mathrm{d-}$ and begins to incr-ease again.* This increase is not due to the oxidation of the primary iron species, but rather results from the contribution of the reduction of protons (5.4) or oxidation of water (5.5) to the total cell current.

Before we perform the third Thought Experiment we need to look more closely at the observations made during the first two, and at the nature and basic properties of interfaces.

5.3 Nature of Interfaces

In the most general definition, an interface is the boundary region between two phases of distinctly different physical properties. Thus, air/water, platinum/elec-trolyte, and copper/gold are examples of different interfaces. Those that involve liquids and mobile charges fall clearly into the domain of electrochemistry. Others involving, for example, two semiconductors are in the area of solid-state physics. Interfaces between insulators and biological tissues could be of interest to bioengi-neers. Interestingly, almost all kinds of interfaces are involved in one or another type of electrochemical sensor.

Curves A and B (Fig. 5.1) describe the behavior of two interfaces that are funda-mentally different. As we have mentioned already, Curve A below the breakdown voltage shows the typical response of a pure capacitor. Thought Experiments I and II also show that this capacitor is located at the interface between the electrode and the ionic sample. Any interface involving mobile charges always separates these charges. In other words, a capacitor forms spontaneously at such interfaces. Because

a minimum of two layers of charge is formed, corresponding to "two plates" of the capacitor, it is customary to call it a double-layer capacitor, and use C_{dl} for its symbol. Charge on at least one plate is distributed in space, forming a so-called space charge. It is described by the Poisson–Boltzmann distribution law, which is a physical expression of "compromise" between Coulombic and kinetic (thermal) interactions between ions. Its thickness depends on the electric field, conductivity of the interfacial materials, and on temperature. Space charge is a common feature not only at electrochemical interfaces but also in semiconductor junctions, in electrolyte solutions, corona discharges, and many other physical scenarios.

Electrochemists have studied metal/electrolyte and semiconductor/electrolyte interfaces for decades (Bockris and Reddy, 1973; Koryta et al., 1993; Bard and Faulkner, 2001). From the electrochemical point of view, the critical aspect is whether these interfaces block the charge transfer. If the interface completely blocks transfer of charge it is called an ideally polarized interface. In other words, the R_{ct} is infinitely large. A schematic of an ideally polarized interface between a metal and an inert electrolyte in which the space charge is formed is shown in Fig. 5.2a, along with its equivalent electrical circuit elements. A mercury electrode immersed in a solution of sodium fluoride (Hg/NaF) comes close to the realization of such an interface. A platinum electrode immersed in a solution of sulfuric acid (Pt/H_2SO_4), which was used to generate Curve A in Fig. 5.1, is another relatively good example of a polarized interface. There are a few other combinations that approach this "ideality," but not many.

Let us now look at the conversion of Curve A to Curve B. What has happened there is that a small charge-transfer resistor has been added in parallel to the double-layer capacitor, through which electrons can shuttle between Pt and the redox couple (5.6). In other words, the addition of the redox couple has converted the polarized interface (Fig. 5.2a) to a nonpolarized interface (Fig. 5.2b).
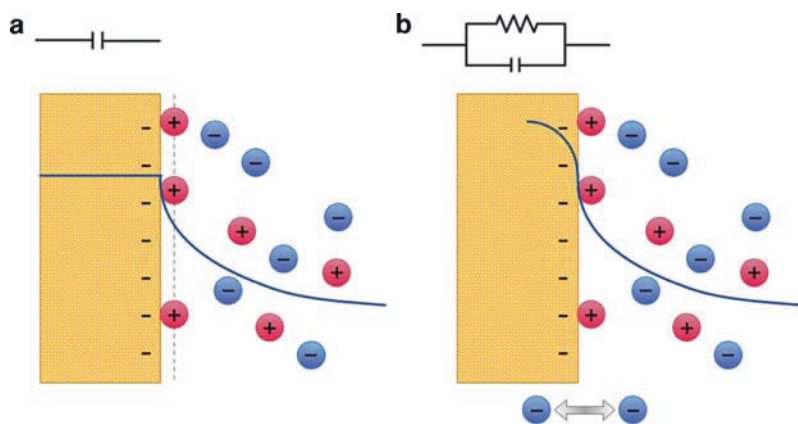

Fig. 5.2 Model of (**a**) polarized and (**b**) nonpolarized interface. The solid line represents potential profile. The dotted line shows the closest approach of adsorbed solvated cations. The solid phase in both cases is electronically conducting

At this interface, charges are separated and form the double-layer capacitor, but because electrons can transfer freely between the two phases, the interfacial charge Q_i is fixed at only one value by the equilibrium potential E_{eq}.

$$Q_i = C_{dl}E_{eq} \tag{5.7}$$

In other words, at the nonpolarized interface, the interfacial potential E_{eq} is uniquely tied by the Nernst equation (5.8) to the activity a_i of the charged species crossing the interface. This is the key relationship in potentiometric sensors (Chapter 6).

$$E = E^0 + \frac{\Re T}{nF}\ln a_i \tag{5.8}$$

On the other hand, equilibrium at the polarized interface is described by the Gibbs–Lippmann equation (5.9). Here, the equilibrium potential E_{eq}, surface concentration $\sum_i \Gamma_i$ of all adsorbing species, their bulk electrochemical potential $\tilde{\mu}_i$, and the resulting interfacial charge Q_i are linked rather less explicitly to surface tension γ.

$$-\gamma = Q_i E + \sum_i \Gamma_i \tilde{\mu}_i \tag{5.9}$$

A notable difference between these two relationships is that the Gibbs–Lippmann equation contains one more independent variable parameter, the interfacial charge. It cannot be determined directly. Several unsuccessful attempts to design chemical sensors (e.g., the immunosensor) based on the measurement of adsorbed surface charge have been made. There are no ideally polarized interfaces that are sufficiently "ideal" to allow such direct measurement of interfacial charge.

Another interface that needs to be mentioned in the context of polarized interfaces is the interface between the insulator and the electrolyte. It has been proposed as a means for realization of "adsorption-based potentiometric sensors" using Teflon, polyethylene, and other hydrophobic polymers of low dielectric constant D_2, which can serve as the substrates for immobilized charged biomolecules. This type of interface happens also to be the largest area interface on this planet: the interface between air (insulator) and sea water (electrolyte). This interface behaves differently from the one found in a typical metal–electrolyte electrode. When an ion approaches such an interface from an aqueous solution (dielectric constant D_1) an image charge is formed in the insulator. In other words, the interface acts as an "electrostatic mirror." The two charges repel each other, due to the low dielectric constant (Williams, 1975). This repulsion is called the Born repulsion H, and it is given by (5.10).

$$H = \left(\frac{Q^2}{4D_1 r}\right)\frac{D_1 - D_2}{D_1 + D_2} \tag{5.10}$$

Thus, both cations and anions are forced away from the insulator/electrolyte interface over a distance r. As a result, their concentration profile is inverted, relative to the one created at the metal/electrolyte interface, as shown in Fig. 5.3. There the ions are attracted to and repelled from the metal surface by Coulombic interactions
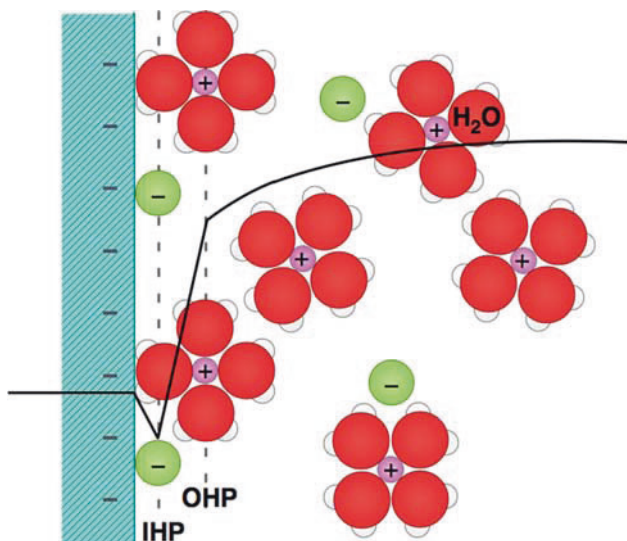
Fig. 5.3 Structure of interface at equilibrium. IHP is the inner Helmholz plane through the contact adsorbed, nonsolvated anions and OHP is the outer Helmholz plane

originating from the potential applied to the electrode. The extent of these interactions depends on the charge on the metal, on the dielectric constant of the solution, and on the ionic strength.

5.4 Nature of the Current–Voltage Curve

We have, at this point, examined some rather complex features of the current–voltage (i–V) curve. As you are now aware, it is affected by the nature and concentration of the electroactive species, by the electrode material, and by the mode of mass transport. The general equation that describes the current density j as the function of applied voltage E is

$$j = nFk_0 \left\{ C_O(0,t)e^{-\alpha f(E-E_0)} - C_R(0,t)e^{(1-\alpha)f(E-E_0)} \right\} \tag{5.11}$$

where $C_O(0,t)$ and $C_R(0,t)$ are the concentrations of the oxidized and of the reduced forms of the redox couple at the surface and at time t, and k_0 is the heterogeneous rate constant. Symbol f is an abbreviation of the multiplier containing a number of electrons n, gas constant $\Re$, absolute temperature T, and the Faraday constant F.

$$f = \frac{nF}{\Re T} \tag{5.12}$$

The symmetry coefficient α is important in electrode kinetics, but less important for the operation of electrochemical sensors. The discussion of its meaning can be found in standard electrochemical textbooks (e.g. Koryta et al., 1993). Its value typically lies between 0.3 and 0.7.

There are two distinctly different regions in the current–voltage curve that are discussed separately in the following Thought Experiment III. First is the charge-transfer region, which lies close to the equilibrium potential. The current densities in this region are very small, so that bulk electrolysis can be neglected. Second is the mass transport region, in which the electrolysis and mass transport dominate the current behavior. From the sensing point of view, these two regions define the operation of potentiometric (Chapter 6) and amperometric (Chapter 7) electrochemical sensors. It is evident that the behavior of all electrochemical interfaces depends largely on the value of the charge-transfer resistor. When R_{ct} has any finite value, the double-layer capacitor is "leaky". This situation corresponds to most "real-life" interfaces.

5.4.1 Throught Experiment III

Let's now increase the sensitivity on the current axis such that Curve A appears to cross the zero current line at only one point. Now we are looking specifically at the round insert area of Fig. 5.1. We label the two curves recorded at this magnification as A$'$ and B$'$, but otherwise everything is the same as in the first experiment.

The crossing point for Curve B$'$ also corresponds to the equilibrium because no net current flows through the interface. Therefore, it is possible to write (5.6) as two equally fast reactions.

$$Fe^{3+} + e \xrightarrow{k_c} Fe^{2+} \tag{5.6a}$$

and

$$Fe^{2+} \xrightarrow{k_a} Fe^{3+} + e^- \tag{5.6b}$$

so that there is no net change of concentrations $C_{Fe^{2+}}$ and $C_{Fe^{3+}}$.

$$\frac{dC_{Fe^{3+}}}{dt} = \frac{dC_{Fe^{2+}}}{dt} \tag{5.13}$$

The same scale applies also to Curve A$'$ but we note that:

1. The equilibrium potential for A$'$ is different from that for B$'$.
2. The slopes of the i–V curves at the crossing points are very different, indicating "slow" (Curve A$'$) and "fast" (Curve B$'$) equilibrium (see Chapter 1).

These observations allow us to make some interim qualitative conclusions. The reactions shown in (5.6a) and (5.6b), viewed separately, represent separate – but equal in magnitude – reduction and oxidation currents flowing through the interface. They correspond to the hypothetical cathodic i_c and anodic i_a currents belonging to the Curve B$'$ and are shown as dashed curves in the round insert of Fig. 5.1. Only at

the equilibrium potential is their value equal, as indicated by the solid arrows, i_0. Currents at that potential are called exchange currents, because they do not result in net electrolysis, only in "zero-sum exchange" of the Fe^{2+} and Fe^{3+} species.

Strictly speaking, the term "exchange current" applies only to conditions when concentrations of oxidized and reduced species are equal, that is, to standard potential. In this chapter, this requirement has been relaxed, and we use this term to describe equal and opposite currents resulting in net zero current at the interface, that is, at equilibrium under at any concentration.

The same argument can be extended to Curve A′, but here the cathodic and anodic currents and also the exchange currents (not shown) are markedly smaller. Obviously, the magnitude of the exchange current is linked to the slope of the $i-V$ curve at E_{eq} and has the meaning of reciprocal resistance. In Fig. 5.1, we see that this slope is given by

$$\left(\frac{di}{dE}\right)_{E_{eq}} = \frac{1}{R_{ct}} \tag{5.14}$$

Now we need to recall the origin of Curves A′ and B′: that is, go back to Curves A and B. At the normal magnification (Thought Experiment II) they "look like" they belong to "ideally polarized" and "nonpolarized" interfaces, respectively. However, at the higher magnification of our focus now, in Experiment III, they reveal that they are both nonpolarized, that is, that the R_{ct} is not infinitely large. An infinitely large R_{ct} Curve A would correspond to the ideally polarized interface and i_0 would be zero.

The impedance of the electrode is inversely proportional to its area A. Therefore, it is the current density j and not the current i that is the dominating factor in the response of the electrochemical cell. Current and current density are simply related by (5.15).

$$j_0 = \frac{i_0}{A} \tag{5.15}$$

Here also lies the reason for making the area of the working electrode much smaller than that of the auxiliary electrode. Because the two electrodes are connected in series, the larger impedance of the two dominates the overall $i-V$ response of the cell. Because we want all the information to originate only from the working electrode, we have to make its area smaller. This point is frequently neglected when, for convenience of fabrication, electrodes of equal area are used in some microfabricated amperometric and conductometric sensors.

5.4.2 Charge-Transfer Region

Charge-transfer resistance and the related exchange current density are the two most important factors in the operation of all electrochemical sensors. They play a role in selectivities, response times, baseline drift, and so on. In the following section, we take a closer look at what they are and how they are determined.

The electrochemical conversions due to the cathodic and anodic exchange current densities are

$$j_c = k_c nF C_O(0) \tag{5.16a}$$

and

$$j_a = -k_a nF C_R(0) \tag{5.16b}$$

The rate constants k_c and k_a for these two processes increase exponentially with the applied potential E.

$$k_c = k_0 \left(e^{-\alpha f(E-E_{eq})} \right) \tag{5.17a}$$

and

$$k_a = k_0 \left(e^{(1-\alpha)f(E-E_{eq})} \right) \tag{5.17b}$$

Consequently, as the applied potential E departs from the equilibrium potential E_{eq} the oxidation or the reduction current increases and dominates the overall reaction.

We can now examine the general properties of (5.11). When the concentrations of **O** and **R** are equal, the equilibrium potential E_{eq} becomes the standard potential E_0. It is a thermodynamic value and a characteristic property of the charge-transferring species. Moreover, because at equilibrium the net current is zero, there is no net electrochemical reaction and time does not play any role. Exchange current densities are also equal, therefore

$$j_c = j_a = j_0 = nF k_0 C \tag{5.18}$$

The current–voltage curve (5.11) is then transformed to the Buttler–Volmer equation.

$$j = nF k_0 C \left(e^{-\alpha f(E-E_0)} - e^{(1-\alpha)f(E-E_0)} \right) \tag{5.19}$$

The major difference between (5.11) and (5.19) is that there are no individual surface concentrations, and the individual rate constants k_a and k_c have been replaced by one heterogeneous rate constant k_0. The price paid for this simplification is that it is valid only in the charge-transfer region of the $i-V$ curve, that is, close to $j = 0$ and at the standard conditions $E_{eq} = E_0$. Nevertheless, it is a stepping stone towards understanding some key parameters of electrochemical sensors.

At equilibrium ($E = E_{eq}$), the net current is zero by definition and j_c equals j_a. Because there is no electrolysis, the surface concentrations $C_O(0)$ and $C_R(0)$ are equal to the bulk concentrations C_O^* and C_R^*, respectively. From (5.11), we have

$$\frac{C_O^*}{C_R^*} = e^{f(E-E_0)} \tag{5.20a}$$

or

$$E = E_0 + \frac{RT}{nF} \ln \frac{C_O^*}{C_R^*} \tag{5.20b}$$

This is the Nernst equation defined from the electrode kinetics considerations. Later, we derive the same relationship on purely thermodynamic grounds.

For values of E close to the standard potential E_0, the exponential terms in (5.19) can be expanded according to the formula $e^x \sim 1 + x$. After substitution from (5.16), we obtain

$$j = j_0 f(E - E_0) \tag{5.21}$$

Differentiation of (5.21) with respect to E and substitution from (5.12) yields

$$\frac{\mathrm{d}j}{\mathrm{d}E} = \frac{j_0 nF}{\Re T} \tag{5.22}$$

which is the slope of the current–voltage curve at E_{eq}. Its inverse has the dimensions of resistance and is, indeed, the earlier-identified, charge-transfer resistance R_{ct}.

$$\frac{\mathrm{d}E}{\mathrm{d}j} = R_{ct} = \frac{\Re T}{nF j_0} \tag{5.23}$$

This means that charge-transfer resistance of fast reversible charge-transfer reactions is low and their exchange current density is high.

5.4.3 Mass Transport Region

When the applied potential is far from the equilibrium potential, current flows through the cell and electrochemical processes take place at both electrodes. Inevitably, the concentrations of the electroactive species in the bulk of the electrolyte are different from those at the electrode surface. Concentration gradients develop and mass transport to and from the electrodes takes place. Therefore, the current through the cell is described by (5.11), rather than by (5.19). As we have seen in Fig. 5.1, current levels off, becoming mass transport limited. The parameter responsible for this limitation is the mass transport resistance. The exact magnitude and time-dependence of the mass transport limited current depends on geometry of the electrodes, duration of the experiment, and on the modes of mass transport, in other words, on the actual values of $C_O(0,t)$ and $C_R(0,t)$. Because it is almost always possible to apply a sufficiently high potential (positive or negative) to the working electrode, the electrode kinetics always becomes fast and the mass transport ultimately becomes the limiting factor. This is clearly seen in Fig. 5.4 where the mass transport regime is reached even for the very slow kinetic processes (e.g., $j_0/j_L = 0.01\,\mathrm{cm\ s^{-1}}$). Thus, kinetics rarely become a limiting factor for the operation of amperometric sensors, but is crucially important for some potentiometric sensors. We should consider these statements in view of Fig. 5.1 and the Thought Experiments I–III. The slow charge-transfer kinetics is desirable because it opens the "electrochemical window" (Curve A, in Fig. 5.1), where we can study relevant electrochemical processes. On the other hand, we generally want this kinetics to be fast for species that we want to determine (the analyte). (The mass transport
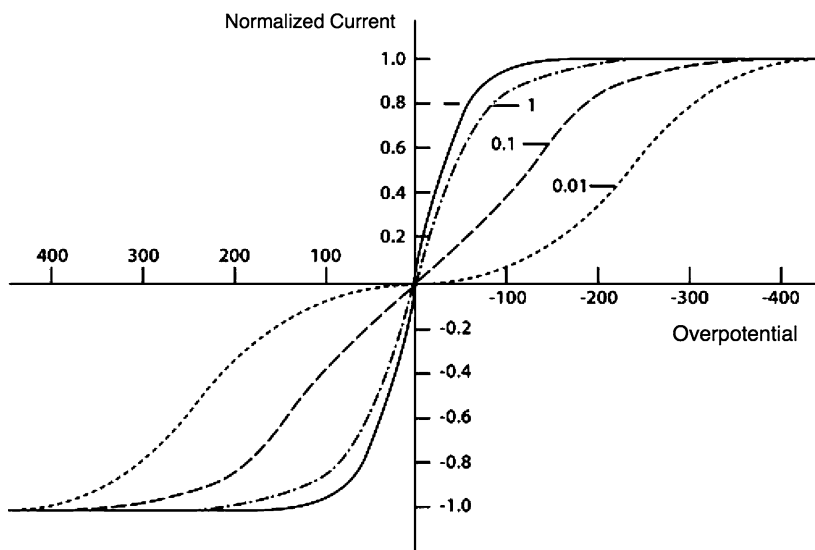
Fig. 5.4 Effect of the exchange current density on the shape of the current–voltage curves normalized to the limiting current I_L (adapted from Bard and Faulkner, 2001, p. 107)

regime is key to the operation of amperometric sensors; it is discussed in Chapter 7.) Here it suffices to say that resistances of both the charge-transfer resistance region and of the mass transport region contribute to the highly nonlinear shape of the current–voltage curve.

5.5 Determination of Cell Parameters: Experimental Methods

Rational optimization of performance should be the main goal in development of any chemical sensor. In order to do that, we must have some quantitative tools of determination of key performance parameters. As we have seen already, for electrochemical sensors those parameters are the charge-transfer resistance and the double-layer capacitance. Particularly the former plays a critical role. Here we outline two approaches: the Tafel plots, which are simple, inexpensive, but with limited applicability, and the Electrochemical Impedance Spectroscopy (EIS), based on the equivalent electrical circuit model, which is more universal, more accurate, and has a greater didactic value.

5.5.1 Tafel Plots

They are obtained in two steps from the Buttler–Volmer equation (5.19) combined with (5.18). The B–V equation is valid only for current densities close to $j = 0$. In
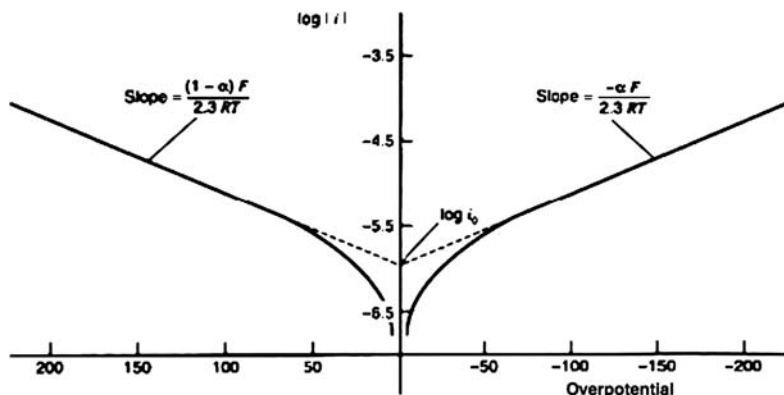
Fig. 5.5 Tafel plots for anodic and cathodic branches of the current-overpotential curve (adapted from Bard, 2001, p. 103)

other words, there can be no appreciable mass transport polarization. For $E - E_0 < 0$, that is, for the "cathodic branch," the second exponential term in (5.19) can be neglected against the first. In other words, the current is dominated only by the reduction process. Taking the base-10 logarithm and rearranging, we obtain (5.24a)

$$\log j_c = \log j_0 - \frac{\alpha n F}{2.3 \Re T}(E - E_0) \tag{5.24a}$$

A similar procedure is followed for $E - E_0 > 0$ which yields

$$\log j_a = \log j_0 - \frac{(1 - \alpha) n F}{2.3 \Re T}(E - E_0) \tag{5.24b}$$

The plot of $\log j$ versus $E - E_0$ is known as a Tafel plot (Fig. 5.5). The two branches intercept at $\log j_0$ from which R_{ct} can be determined according to (5.23). The plot should be symmetrical about the vertical axis for $\alpha = 0.5$, which can be determined from either slope.

Tafel plots have been used successfully for evaluation of slow electrochemical reactions at metal electrodes. Their application to electrochemical sensors is somewhat limited because of the mass transport boundary condition imposed by the nature of the Buttler–Volmer equation. Nevertheless, because it is simple and inexpensive, it should be always tried as the first approach, but bearing in mind its limitations.

5.5.2 The Equivalent Electrical Circuit Approach

The equivalent electrical circuit approach has already been introduced in connection with analysis of mass sensors (Chapter 4). Its application is older and somewhat

simpler in the analysis of electrochemical systems. It has been used extensively in analysis of ion-selective electrodes (Demoz et al., 1995) and electrode kinetics (Macdonald, 1987; Bard and Faulkner, 2001, Ch. 10). Because frequency is an essential parameter, the method is often called electrochemical impedance spectroscopy. The absence of inductance in most electrochemical processes makes its application considerably simpler than the equivalent circuit analysis used for QCM. The second advantage comes from the fact that the model made of equivalent resistors and capacitors mirrors the same elements found in electrochemical processes. Thus, the equivalency becomes even more intuitive.

The equivalent circuit representing the electrochemical cell in Fig. 5.1 (insert) is shown in Fig. 5.6a. Its overall impedance Z_{cell} is

$$Z_{cell}(\omega) = Z_W(\omega) + R_s + Z_{AUX}(\omega) \tag{5.25}$$

Because there is no charge separation (and hence no capacitor) involved in the bulk electrolyte, electrolyte resistance R_s is written as normal resistance, which does not have any frequency dependence. Because impedances of the working and the auxiliary electrodes are inversely proportional to their area, by choosing $A_{AUX} >> A_W$, the contribution to the cell impedance of the auxiliary electrode can be neglected. The argument for having a large auxiliary electrodehas been phrased differently
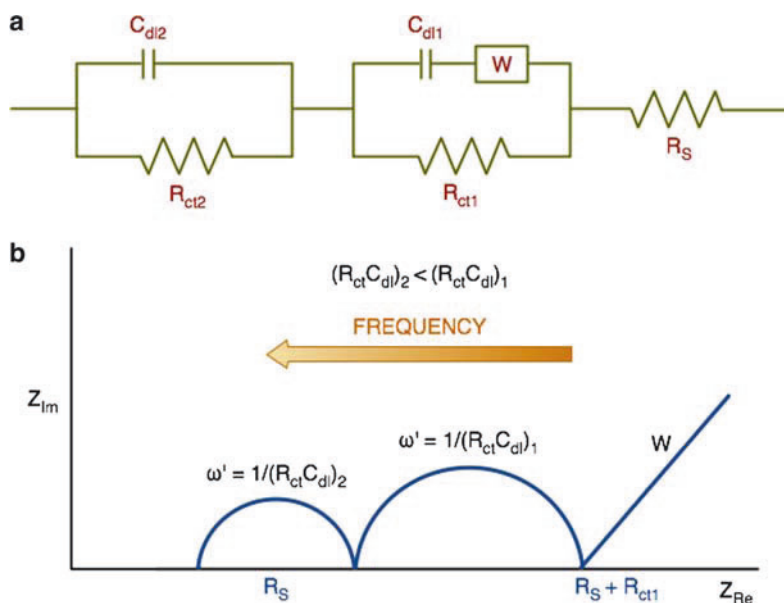

Fig. 5.6 Equivalent electrical circuit of electrochemical cell (*top*) and corresponding Nyquist plot containing Warburg impedance *W* (*bottom*)

earlier (5.15), but its physical meaning is the same. The cell impedance can then be written as a complex number.

$$Z_{cell}(\omega) = R_S + Z_W(\omega) = R_S + \frac{R_{ct}}{1 + j_{Im}\omega R_{ct}C_{dl}} \tag{5.26}$$

The symbol ω is angular frequency. Note that j_{Im} denotes an imaginary number, and not a symbol for current density. Next, we perform the separation of the real and the imaginary parts by multiplying the numerator and the denominator of the second term of (5.26) with $(j_{Im}\ \omega R_{ct}C_{dl}-1)$. Thus, we obtain for the cell impedance:

$$Z_{cell}(\omega) = R_S + \frac{R_{ct}}{1 + (\omega R_{ct}C_{dl})^2} - j_{Im}\omega\frac{R_{ct}^2C_{dl}}{1 + (\omega R_{ct}C_{dl})^2} \tag{5.27}$$

In (5.27), the second term is the real part of the electrode impedance Z_{Re} and the third term is its imaginary part Z_{Im}.

$$Z_{Re} = \frac{R_{ct}}{1 + (\omega R_{ct}C_{dl})^2} \tag{5.28a}$$

and

$$Z_{Im} = j\omega\frac{R_{ct}^2C_{dl}}{1 + (\omega R_{ct}C_{dl})^2} \tag{5.28b}$$

A plot of Z_{Im} against Z_{RE}, called the Nyquist plot, is a semicircle centered on the real axis (Fig. 5.6b). In this plot, the angular frequency ω is the parameter that increases in the anticlockwise direction. We can now perform a simple phenomenological analysis of the plot. From (5.27), we see that at low frequencies (i.e., at "almost DC" behavior)

$$\lim_{\omega \to 0} Z_{cell} = R_S + R_{ct} \tag{5.29}$$

On the other hand, for very high frequencies, the electrolyte resistance R_S dominates. That is, by the way, the principal reason for using high-frequency excitation in conductometric sensors (Chapter 8) when we want to avoid polarization of the electrodes.

$$\lim_{\omega \to \infty} Z_{cell} = R_S \tag{5.30}$$

The frequency ω_τ at which the values of the real and imaginary components are equal has special meaning. Thus from (5.28), when $Z_{Re} = Z_{Im}$,

$$\omega_\tau = \frac{1}{(R_{ct}C_{dl})_\tau} \tag{5.31}$$

The time constant τ of the cell is then

$$\tau = \frac{2\pi}{\omega_\tau} = (R_{ct}C_{dl})_\tau \tag{5.32}$$

Equations (5.29) through (5.32) contain all the information about the electrochemical cell. When multiple interfaces are involved in the cell, in parallel or in series

combinations, it is possible to use the conversion formulas shown in Appendix D. The electrochemical impedance spectroscopy is therefore one of the most versatile and powerful tools of electrochemical investigations. It should be noted that for electrochemical cells in which various impedances in series have comparable value, the Nyquist plot analysis yields multiple semicircles corresponding to such impedances. Even when the semicircles begin to overlap it is often possible to deconvolute them, and to obtain individual impedances.

Food for Thought #5

Electrochemical Divider

1. The equivalent circuit of an electrochemical cell is shown in Fig. 5.6. It can be represented by a capacitive divider consisting of C_W and C_{AUX}, connected in series. Figure out how the voltage V and charge Q are distributed across this divider when the resistances are (a) finite (b) infinite.

Electrochemical Immunosensors

Antibodies, like all proteins, have a charge at all pH values except one (the isoelectric point). When they combine specifically with the corresponding antigen this charge may, and usually does, change. It has been suggested that when this reaction takes place at the interface between the electrolyte and a polarized electrode, the change of charge could be measured directly.

2. For a DC measurement, what type of interface would that have to be, ideally polarized or nonpolarized? Would the same restriction apply in the case of an AC measurement?

Interfaces

It has been said that interfaces separate charges. Explain the charge separation and draw the equivalent circuit diagram of the following materials interfacing with an electrolyte containing only NaCl and water.

3. Air
4. Platinum
5. Glass
6. Polyethylene

Mixed Potential

7. In the round insert of Fig. 5.1, we see that at least two equilibrium potentials exist in the aqueous solution of any redox couple. Yet there can be only one potential at any given electrode. How do you explain this apparent discrepancy?

Nyquist Plots

In the impedance circuit analysis of an electrode/solution interface, the impedance and phase angle (or real and imaginary components) are measured at several frequencies. The imaginary component is plotted against the real component and the semicircle is then interpolated to those points. The solution resistance and the sum of solution and charge-transfer resistances, respectively, are then found by extrapolating the semicircle to $Z_{Im} = 0$.

8. If the objective of this exercise is to find charge-transfer resistance, why is it not possible to use a DC current (i.e., $\omega = 0$) in a concentrated solution of electrolyte, where R_s could be neglected?

Voltage and Potential

9. Voltage is the difference between two potentials. Consider capacitors C_W and C_{AUX} of the working and of the auxiliary electrode. What are the potentials with respect to the solution of the small working electrode and of the large auxiliary electrode in the experiment described in Curve A (Fig. 5.1) (a) before any voltage is applied to the cell (i.e., $V = 0$) and (b) after a voltage $V = +2.0\,V$ is applied?

Symbols

a_i	Activity
A	Area
C	Molar concentration
C_{dl}	Double-layer capacitance
D	Dielectric constant
E	Potential
F	Faraday constant (96,493 Coulombs)
H	Born repulsion
I	(or i) current

j	Current density
j_{Im}	An imaginary number
k	Rate constant
m	Number of moles
n	Number of electrons
r	Distance
R	Resistance
V_{cell}	Volume of the cell
Z	Impedance
α	Symmetry coefficient
τ	Time constant
$\tilde{\mu}_i$	Bulk electrochemical potential
γ	Surface tension
$\sum_i \Gamma_i$	Summation of surface concentrations of all species i
Q	Charge
$\Re$	Gas constant ($8.314\,\mathrm{J\,K^{-1}\,mol^{-1}}$)

References

Bard, A.J. and Faulkner, L.R. (2001) *Electrochemical Methods, Fundamentals and Applications*, 2nd ed. Wiley.

Bockris, J.O'M. and Reddy, A.K.N. (1973) *Modern Electrochemistry*. Plenum.

Demoz, A., Verpoorte, E.M.J., and Harrison, D.J. (1995) *J. Electroanal. Chem.* 71–78.

Koryta, J., Dvorak, J., and Kavan, L. (1993) *Principles of Electrochemistry*, 2nd ed. Wiley.

Macdonald, J.R. (1987) *Impedance Spectroscopy*. Wiley/Interscience.

Williams, R. (1975) *J. Phys. Chem.* 79, 1274–1276.

Chapter 6
Potentiometric Sensors

6.1 Introduction

Potentiometric measurements are done under the condition of zero current. There-
fore, the domain of this group of sensors lies at the zero-current axis (see Fig. 5.1).
From the viewpoint of charge transfer, there are two types of electrochemical inter-
faces: ideally polarized (purely capacitive) and nonpolarized. As the name implies,
the ideally polarized interface is only hypothetical. Although possible in princi-
ple, there are no chemical sensors based on a polarized interface at present and we
consider only the nonpolarized interface at which at least one charged species parti-
tions between the two phases. The "Thought Experiments" constructed in Chapter 5,
around Fig. 5.1, involved a redox couple, for the sake of simplicity. Thus, an elec-
tron was the charged species that communicated between the two phases. In this
section and in the area of potentiometric sensors, we consider any charged species:
electrons, ions, or both.

6.2 Ion Sensors

6.2.1 Interfaces in Ion Sensors

As we have seen already, at zero net current, the current–voltage equation simplifies
to (5.8), which is a form of the Nernst equation, the basic relationship governing
the operation of most potentiometric ion sensors. However, partitioning of only one
type of charged species across the interface is a rather unlikely situation. Usually
there are several charged species that can cross the interface under the given condi-
tions. Partial exchange current densities can be assigned to every one of them. The
relative magnitude of those partial exchange current densities is closely related to
the selectivity of the interface. The interface is said to be selective to the species
with the highest partial exchange current density (Cammann, 1985). According to

the number of charged species that are involved in the overall charge transfer, we classify the nonpolarized interfaces as

(A) Ion-selective: only one type of ion can exchange
(B) Semipermeable: one type of ion cannot exchange
(C) Nonselective: all ions can easily exchange (low R_{ct})
(D) Nonselective: all ions can poorly exchange (high R_{ct})

In all four cases, the condition of the equality of the electrochemical potential $\tilde{\mu}_i$ of the charged species that can cross the interface must hold. The condition of electroneutrality must also apply to all thick electrically conducting phases at equilibrium and is equivalent to the statement that there is no electric field (potential gradient) in the bulk of a conducting phase at zero current. In order to satisfy this condition, the phase must be thicker than the space charge. The thickness of the space charge depends on temperature, on the dielectric constant and on conductivity of the membrane. The ion-selective interface and the nonselective interface with low value of charge-transfer resistance are typical components of potentiometric ion sensors. The semipermeable and nonselective (high R_{ct}) interfaces are common sources of problems in potentiometry.

6.2.1.1 Ion-Selective Interface

This interface is also known as the perm-selective interface (Fig. 6.1a). It is found in ion-selective sensors, such as ion-selective electrodes and ion-selective field-effect transistors. It is the site of the Nernst potential, which we now derive from the thermodynamic point of view. Because the zero-current axis in Fig. 5.1 represents the electrochemical cell at equilibrium, the partitioning of charged species between the two phases is described by the Gibbs equation (A.20), from which it follows that the electrochemical potential of the species i in the sample phase (S) and in the electrode phase (m) must be equal.
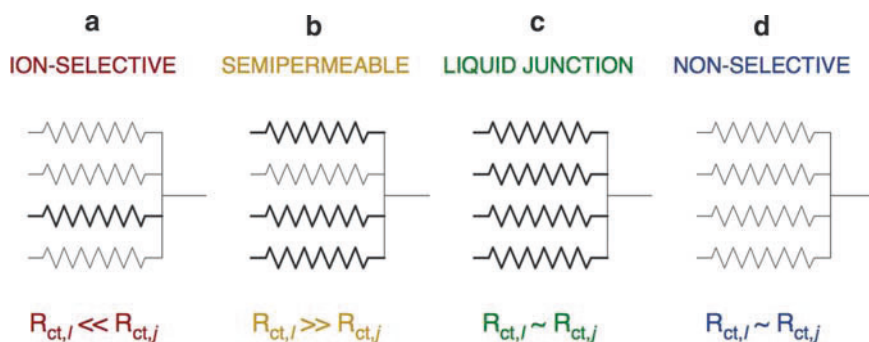

Fig. 6.1 Types of nonpolarized interfaces and corresponding potentials. The bold lines represent transferring charges that have high exchange current density (i.e., low charge-transfer resistance) and dominate the interfacial potential

$$\tilde{\mu}_i^m = \tilde{\mu}_i^S \tag{6.1}$$

The electrochemical potential of the charged species i in a given phase contains terms originating from both the electrical and the chemical work. Thus, for the electrochemical potential of charged species i in the electrode we can write

$$\tilde{\mu}_i^m = \mu_i^{0,m} + \Re T \ln a_i^m + zF\varphi^m \tag{6.2}$$

and similarly for the solution phase

$$\tilde{\mu}_i^S = \mu_i^{0,S} + \Re T \ln a_i^S + zF\varphi^S \tag{6.3}$$

Each phase is charged and has electrostatic potential φ, which is called the Galvani potential. The profiles of the ion activity, of the Galvani potential, and of the electrochemical potential across the sample/electrode interface are shown in Fig. 6.2.

Let us assume that at time $t = 0$ there is some initial activity of a_0 of the exchanging ion in the solution and that the electrostatic potentials of the two phases are equal. When the equilibrium is established, a small number of cations are transferred from the solution to the electrode phase. This changes the electrostatic potential difference between the two phases, with the electrode becoming more positive with respect to the sample. The interfacial potential π is obtained from (6.1)–(6.3).

$$\pi = \varphi^m - \varphi^S = \frac{\mu_i^{0,S} - \mu_i^{0,m}}{z_i F} + \frac{\Re T}{z_i F} \ln \frac{a_i^S}{a_i^m} \tag{6.4}$$

The first term on the right-hand side of (6.4) is the standard potential π^0.
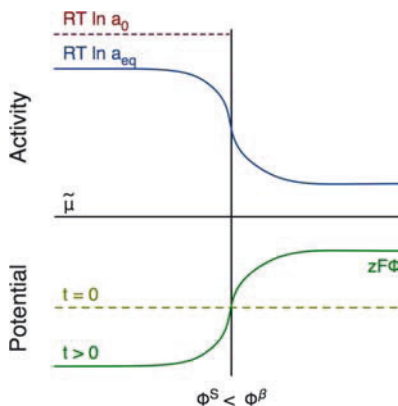

Fig. 6.2 Activity and potential profiles at nonpolarized interface at equilibrium

$$\pi^0 = \frac{\mu_i^{0,S} - \mu_i^{0,m}}{z_i F} \tag{6.5}$$

Thus, the Nernst equation relates the potential difference $\pi = \varphi^\beta - \varphi^s$ at the interface to the activities of species i in phases m and S. The standard state of the metal cation in the metal is unity. In other cases, it is possible to assume that the activity inside the nonmetallic electrode phase is constant. Therefore, the Nernst equation simplifies to

$$\pi = \pi^0 + \frac{\Re T}{z_i F} \ln a_i^S \tag{6.6}$$

For a membrane that is symmetrically placed between two solutions S_1 and S_2 containing the primary ion, such as in a conventional ion-selective electrode with a constant internal reference solution S_2 (Fig. 6.3), the Galvani potentials inside the membrane phase cancel out and (6.6) applies again.

It is important to note that the electrode potential is related to activity and not to concentration. This is because the partitioning equilibria are governed by the chemical (or electrochemical) potentials, which must be expressed in activities. The multiplier in front of the logarithmic term is known as the "Nernst slope". At 25°C it has a value of $59.16\,\mathrm{mV}/z_i$. Why did we switch from n to z when deriving the Nernst equation in thermodynamic terms? Symbol n is typically used for the number of electrons, that is, for redox reactions, whereas symbol z describes the number of charges per ion. Symbol z is more appropriate when we talk about transfer of any charged species, especially ions across the interface, such as in ion-selective potentiometric sensors. For example, consider the redox reaction $Fe^{3+} + e = Fe^{2+}$ at the Pt electrode. Here, the $n = 1$. However, if the ferric ion is transferred to the ion-selective membrane, $z = 3$; for the ferrous ion, $z = 2$.
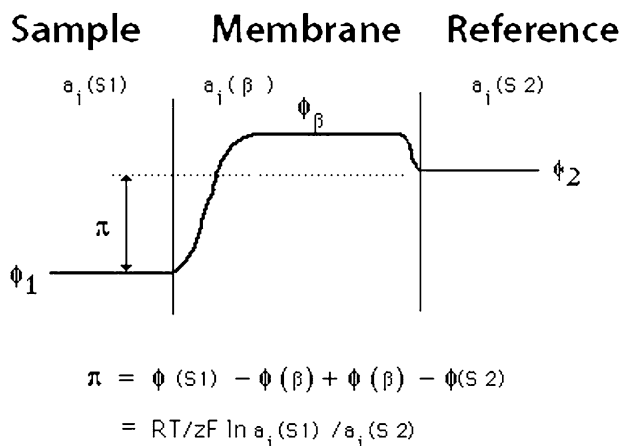

Fig. 6.3 Potential profile through ion-selective membrane symmetrically placed between two solutions

6.2.1.2 Semipermeable Interface

In this case, most ions can pass through the interface but a few are excluded. This situation, which is depicted in Fig. 6.4, is common in dialysis when small ions and low molecular weight substrates can move across the dialysis membrane, but large polyelectrolyte ions, such as proteins, cannot.

This exclusion is based on size of the ion and leads to the formation of the Donnan potential, first mentioned in Chapter 2. Its origin can be explained using the simplest case, involving a uni-univalent electrolyte (NaCl) and a large polyelectrolyte anion $\mathbf{P^{z-}}$, which is present only in the left compartment (marked with a $*$ in Fig. 6.4) and carries z negative charges. We again recall two conditions that must be satisfied at this membrane equilibrium:

1. The equality of the electrochemical potential at the two sides of the membrane for those species that can communicate across it.
2. The condition of electroneutrality, which must exist in both compartments.

The first condition leads to the potential difference shown in (6.3) which, when written for the specific ions in question, is

$$\pi = \varphi^{S_1} - \varphi^m + \varphi^m - \varphi^{S_2} = \frac{\Re T}{zF} \ln \frac{C_{\mathrm{Na}}^*}{C_{\mathrm{Na}}} \tag{6.7}$$

The superscript m is used for the potential of the interior of the membrane, which again cancels out in a symmetrical arrangement. Because only one potential difference can exist across the membrane, the partitioning of the chloride ions is uniquely related to that of the sodium ions.

$$\frac{\Re T}{zF} \ln \frac{C_{\mathrm{Na}}^*}{C_{\mathrm{Na}}} = -\frac{\Re T}{zF} \ln \frac{C_{\mathrm{Cl}}^*}{C_{\mathrm{Cl}}} \tag{6.8}$$

The negative sign in (6.8) is due to the negative charge of the anion. The condition of electroneutrality in each compartment dictates that, on the side of the membrane without polyelectrolytes, the concentration of the salt is

$$C_{\mathrm{Na}} = C_{\mathrm{Cl}} = C \tag{6.9}$$
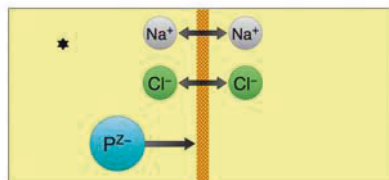
and that, on the other side of the membrane,

Fig. 6.4 Semipermeable interface at which transfer of ion $\mathbf{P^{z-}}$ is blocked

$$zC_P^* + C_{Cl}^* = C_{Na}^* \tag{6.10}$$

From (6.8)–(6.10), we obtain the ratio of ion concentrations in the two compartments separated by the dialysis membrane.

$$\frac{C_{Na}^*}{C_{Na}} = \frac{-zC_P^*}{2C} + \left[1 + \left(\frac{zC_P^*}{2C}\right)^2\right]^{1/2} \tag{6.11}$$

We see that for $C_P^* = 0$, the concentration of sodium ion in both compartments is the same (i.e., $C_{Na} = C_{Na}^*$) and that there is no Donnan potential across the membrane. In other words, it is the presence of the blocked polyelectrolyte that gives rise to the Donnan potential. For low molar concentration of P and high concentration of NaCl, the squared term $(zC_P^*/2C)^2$ can be neglected and (6.11) then changes to

$$\frac{C_{Na}^*}{C_{Na}} = 1 - \frac{zC_P^*}{2C} \tag{6.12}$$

Thus, for a negatively charged polyanion, the ratio $C_{Na}^*/C_{Na} > 1$. This means that sodium ion compensates the excess of the negative charge due to the polyanion, resulting in a potential across the membrane.

$$\pi = \frac{\Re T}{F} \ln\left(1 - \frac{zC_P^*}{2C}\right) \tag{6.13}$$

Even for this simplest of all situations we had to make a fairly drastic assumption of high concentration of NaCl, in order to get from (6.11) to (6.13). The situation is considerably more complicated when different multivalent ions are present in the solution, although the basic argument is the same. In biological fluids, such as whole blood, the value of the Donnan potential across the dialysis membrane can be tens of millivolts. (In the above derivation of the Donnan potential, concentrations instead of activities have been used for purely historical reasons.)

The equivalent circuit corresponding to this interface is shown in Fig. 6.1b. The charge-transfer resistances for the exchange of sodium and chloride ions are very low, but the charge-transfer resistance for the polyanion is infinitely high. There is no direct sensing application for this type of interface. However, it is relevant for the entire electrochemical cell and to many practical potentiometric measurements. Thus if we want to measure the activity of an ion with the ion-selective electrode it must be placed in the same compartment as the reference electrode. Otherwise, the Donnan potential across the membrane will appear in the cell voltage and will distort the overall result.

6.2.1.3 Nonselective Interface with Low R_{ct}

This interface is typical for liquid junctions. This important element of an electrochemical cell is a form of ionic contact that physically separates two solutions,

typically the reference electrode compartment and the sample. It is not an "interface" in the usual meaning of that word because the phases that it separates may have the same or very similar physical properties. However, their chemical composition is different. Also, from the geometrical point of view it is not a sharp boundary. It can be a narrow channel filled with the same solvent, but separating two solutions of different composition. There are concentration gradients of all the species (including the solvent), and therefore the gradients of the chemical potentials of all those species in such a channel. However, because the channel physically separates the two compartments, this arrangement can be viewed also as a "completely nonselective" interface. The exchange current densities of all ions in the liquid junction are high (i.e., R_{ct} is low). The equivalent schematic is shown in Fig. 6.1c.

The liquid junction is a practical necessity in most electrochemical measurements and in almost all potentiometric measurements. It is a site of the dreaded liquid junction (or diffusion) potential E_j. Because this potential is the result of the diffusion, which is a nonequilibrium process, any potentiometric measurement, which uses a liquid junction is a "nonequilibrium" measurement, by definition. The origin of E_j is explained in Fig. 6.5.

Let us assume that a narrow capillary separates two solutions of NaCl in which the relative concentrations are $C_A > C_B$. Thus, a concentration gradient of NaCl exists along the capillary, and Na^+ and Cl^- ions move along this gradient. Therefore, as the diffusion (driven by the chemical potential gradient) takes place, faster chloride ions run ahead of the slower Na^+, thus causing separation of charge and creation of an electric field in the junction. However, this field counteracts further separation of the two ions and, after a short time, steady-state is reached at which the cations and the anions move with the same velocity. This is called a coupled diffusion/migration. It is described by the Nernst–Planck equation, in which for the sake of simplicity the ion flux J_i is considered only in direction x and without convection.

$$J_i = -D_i \frac{da_i}{dx} - \frac{z_i F D_i a_i}{\Re T} \frac{dE_j}{dx} \tag{6.14}$$

Here D_i is the diffusion coefficient of species i, and E_j is the liquid junction potential that develops along the channel. The electrolytic mobility u_i is the limiting velocity of an ion in the electric field of unit strength. It has dimensions of $cm^2\,s^{-1}\,V^{-1}$ and
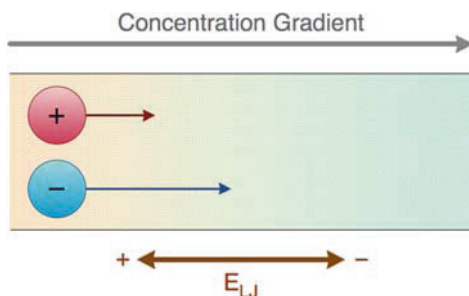

Fig. 6.5 Origin of the liquid junction potential in an "open capillary junction"

is related to the diffusion coefficient by (6.15).

$$u_i = D_i \frac{|z_i|F}{\Re T} \tag{6.15}$$

In the example of the NaCl solution, the mobility of Cl^- is higher ($u_{Cl} = 7.91 \times 10^{-4}\,cm^2\,s^{-1}\,V^{-1}$) than that of Na^+ ($u_{Na} = 5.193 \times 10^{-4}\,cm^2\,s^{-1}\,V^{-1}$).

The diffusion is driven by the gradient of the chemical potential which, in turn, is caused by the activity gradient da_i/dx. The potential gradient (electric field) dE_j/dx acting on the ions generates an electrostatic force that causes their migration (the second term in (6.14). In a liquid junction, the electric field created by separation of ions of different mobility then acts against the diffusion gradient, resulting in formation of steady-state and constant E_j. Under those conditions, steady-state is reached when the change of the electrochemical potential of every species is zero.

The conductivity κ is an intrinsic measurable property of the electrolyte solution. It is the sum of contributions to conductivity of all the ions present in the solution

$$\kappa = F \sum_i |z_i| u_i C_i \tag{6.16a}$$

Conductivity divided by concentration and multiplied by factor 1,000 is called "equivalent conductivity."

$$\Lambda = \frac{1000\kappa}{C} \tag{6.16b}$$

The fraction of charge transported by ion i, relative to the total transported charge, is called the transference number t_i of ion i. It is expressed as

$$t_i = \frac{|z_i| u_i C_i}{\sum_j |z_j| u_j C_j} \tag{6.17}$$

The contribution of the individual ion i to the equivalent conductivity Λ of the electrolyte is called equivalent ionic conductivity (λ) where λ = unit conductivity/ concentration. For simple uni-univalent electrolytes, such as Na^+, Cl^-, we can write for the individual contributions of the anion and cation

$$\Lambda = \lambda_+ + \lambda_- \tag{6.18}$$

They are related to the mobility by the Faraday constant.

$$\lambda_i = |z_i| F u_i \tag{6.19}$$

In general, for mixed valency electrolytes, we can express the individual transference numbers in terms of the experimentally accessible equivalent ionic conductivities from (6.17) and (6.19) as

$$t_i = \frac{|z_i| \lambda_i C_i}{\sum_j |z_j| \lambda_j C_j} \tag{6.20}$$

If the liquid junction is formed between two aqueous solutions of electrolytes containing many types of ions of different valency and at different concentrations, the electrochemical potentials of all species are linked by the Gibbs–Duhem equation ((A.21), Appendix A). For any moving species, the change of its electrochemical potential is caused by the change of its molar free energy $\bar{G}_i$.

$$d\bar{G}_i = \frac{t_i}{z_i} d\tilde{\mu}_i \tag{6.21}$$

Therefore, at steady-state, the integral of all molar free energy contributions must be zero.

$$\int\limits_{x=0}^{x=L} d\bar{G} = 0 = \sum_i \int\limits_{x=0}^{x=L} \frac{t_i}{z_i} d\tilde{\mu}_i \tag{6.22}$$

The liquid junction potential E_j is then obtained by expanding the electrochemical potential $\tilde{\mu}_i$ (A.19b) and integrating (6.22) over the length of the junction L. In this step, we assume that the liquid junction separates two aqueous solutions, and the standard states μ^0 are the same on both sides of the junction. The complete integral describing the junction potential is then

$$E_j = -\frac{\Re T}{F} \sum_i \int\limits_{x=0}^{x=L} \frac{t_i}{z_i} d(\ln C_i + \ln f_i) \tag{6.23}$$

Unfortunately, the spatial functions of the activity coefficient and of the transference numbers in the general junction are not known and as a result (6.22) cannot be integrated. This situation becomes much worse when the junction connects two different solvents. In such cases the resulting measurement error from undefined E_j can reach several hundred millivolts.

The practical way out of this problem is through a special experimental arrangement, which allows the introduction of justifiable simplifying assumptions. The two most important ones are the constancy of the activity coefficient and the linear concentration profiles throughout the junction. The physical implementation of such an arrangement is the open capillary (Henderson junction) between compartments A and B, containing only two uni-univalent electrolytes with one common ion, for example, KCl + NaCl or HCl + HNO$_3$. In that case, the gradients are nearly linear and (6.23) can be integrated to yield (6.24).

$$E_j = \frac{\sum_i \frac{|z_i| u_i}{z_i} [C_i(A) - C_i(B)]}{\sum_i |z_i| u_i [C_i(A) - C_i(B)]} \frac{\Re T}{F} \ln \frac{\sum_i |z_i| u_i C_i(A)}{\sum_i |z_i| u_i C_i(B)} \tag{6.24}$$

For a uni-univalent electrolyte, the multiplier term in (6.24) cancels out and the liquid junction potential can be calculated from the equivalent ionic conductivities λ_i of the two solutions, from the Sargent equation.

$$E_j = \pm \frac{\Re T}{F} \ln \frac{(\lambda_+ + \lambda_-)_A}{(\lambda_+ + \lambda_-)_B} \qquad (6.25)$$

That equation explicitly shows the dominating influence of the ion mobilities on the value of the liquid junction potential. Although its application is limited, the Henderson junction is a significant experimental achievement because it allows the measurement of one electrode against another with known contribution of the liquid junction potential.

In a practical junction, the concentration of the internal solution is usually kept as high as possible, so that the charge transfer at the junction is dominated by the salt that has approximately equal mobilities of the anion and of the cation (i.e., equitransferrant electrolytes). For example, a saturated aqueous solution of KCl ($u_{K+} = 7.619 \times 10^{-4}\,\text{cm}^{-1}\,\text{s}^{-1}\,\text{V}^{-1}$, $u_{Cl-} = 7.912 \times 10^{-4}\,\text{cm}^2\,\text{s}^{-1}\,\text{V}^{-1}$) satisfies this condition. Because the internal salt bridge solution can leak out and contaminate the sample, it is desirable to have other equimobile salts available when necessary. These are, for example, Li acetate (4.01×10^{-4} and 4.24×10^{-4}) and NH_4NO_3 (7.619×10^{-4} and 7.404×10^{-4}). They are used in a configuration known as the *salt bridge*. It is a combination of two liquid junctions connected back to back (Fig. 6.6a). By using a salt bridge, the liquid junction potential can be reduced to less than 1 mV. The double liquid junction arrangement (Fig. 6.6b) is used when contamination of the sample by ions of the internal compartment needs to be avoided.

By far the biggest problems with the stability and the magnitude of the liquid junction potentials arise in applications where the osmotic or hydrostatic pressure, temperature, and/or solvents are different on either side of the junction. For this reason, the use of an aqueous reference electrode in nonaqueous samples should be avoided at all cost because the gradient of the chemical potential of the solvent has a very strong effect on the activity coefficient gradients of the ions. In order to circumvent these problems one should always use a junction containing the same solvent as the sample and the reference electrode compartment.
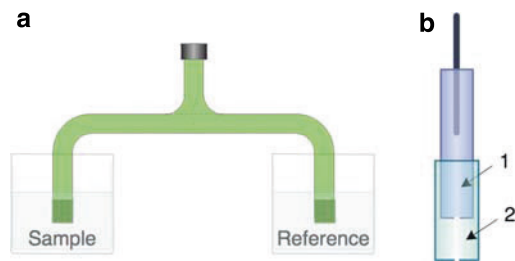

Fig. 6.6 (**a**) Salt bridge and (**b**) double liquid junction. Composition of solutions 1 and 2 can be selected such that the contamination of the sample is prevented
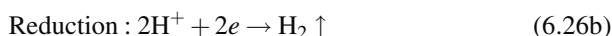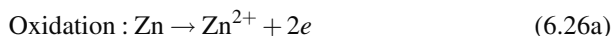
6.2.1.4 Nonselective Interface with High R_{ct}

The last type of interfacial potential that is encountered at some nonpolarized interfaces is the mixed potential (Fig 6.1d). Although the net current flowing through the interface is zero, this is not the equilibrium potential. It can be found at the membranes at which the partial exchange current densities are low and are of the same order of magnitude, for example, at ion-selective membranes that have a very low selectivity. The distinguishing feature of the mixed potential from the liquid junction potential (which is intentionally nonselective) is the magnitude of the sum of the partial current densities. For a mixed potential, they are typically less than 10^{-6} A cm^{-2}. For physical explanation of the origin of this potential we need to refer to the current–voltage equation (5.11), which can be written for any charged species that crosses the interface, including ions. If two different ions are crossing the interface the partial exchange current density can be obtained for each of them. At zero current two conditions apply: there is only one potential at the interface common to all ions, and the sum of all partial currents must be zero.

Even for uni-valent electrolytes, the solution of these equations leads to intractable algebra and some drastic assumptions have to be made in order to obtain an approximate solution (Cammann, 1985). Here we show the origin of the mixed potential graphically (Fig. 6.7).

The movement of ions to and from the membrane is shown as superscript arrows. Because the equilibrium potentials and the exchange current densities for each individual process are different, the mixed potential E_{mix} lies somewhere between the two equilibrium potentials, so that it satisfies the condition of zero net current. If the rates of the individual charge-transfer reaction change, for example, due to some specific adsorption, due to relative motion of the electrolyte with respect to the stationary membrane, or due to uncontrollable changes of the membrane, the E_{mix} also changes. It is important to realize that different ions, and their corresponding exchange current densities, respond differently to these effects. This can explain why a seemingly "selective response" is obtained for certain kinds of adsorptions, such as immunochemically mediated adsorption, provided that the experimental conditions are rigidly maintained.

The mixed potential accounts for a large portion of reported artifacts in the "unorthodox" potentiometric sensors, particularly biosensors, and can be rightfully called "evil potential". The physical origin of such artifacts can be illustrated using a simple example. Let us assume that a multiple electron transfer takes place simultaneously at the interface of a lump of Zn immersed in dilute HCl. Because this metal is not externally connected the net current is zero. The redox reactions taking place are as follows.

$$\text{Oxidation}: \text{Zn} \rightarrow \text{Zn}^{2+} + 2e \tag{6.26a}$$

$$\text{Reduction}: 2\text{H}^+ + 2e \rightarrow \text{H}_2 \uparrow \tag{6.26b}$$

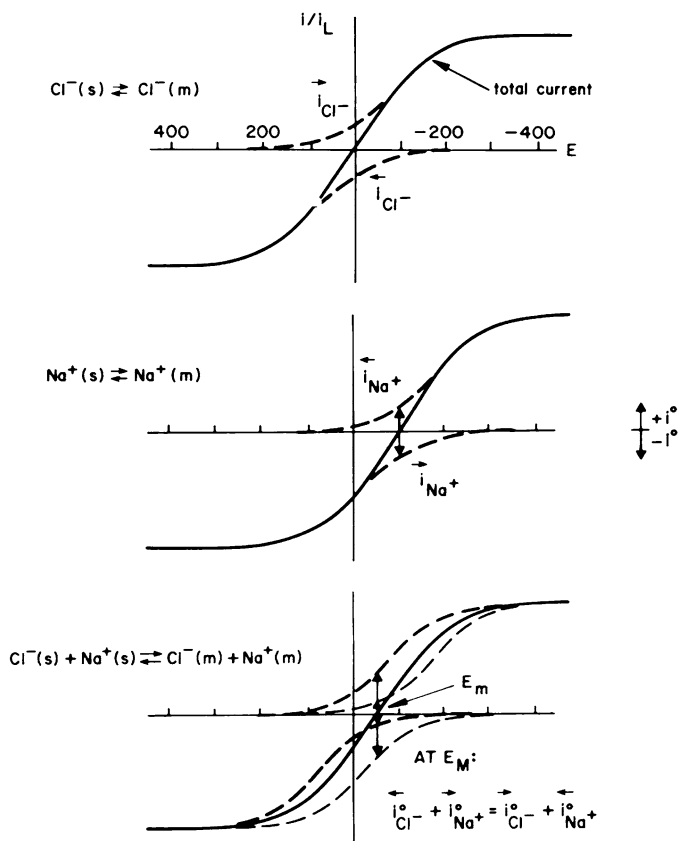In this experiment, the dissolution of Zn is accompanied by the evolution of hydrogen.

Fig. 6.7 Origin of the mixed potential

This is schematically depicted in Fig. (6.8a). The redox processes controlled by (6.26a) and (6.26b) are independent of each other, yet they are bound by the condition that the electrons are internally utilized; that is, the net current is zero. If we now stir the solution or change the gas pressure, the mass transport of Zn^{2+}, H^+, and H_2 to and from the metal interface will change independently, thus changing the ratio of exchange currents i_c and i_a. Consequently, E_{mix} will change accordingly. (Note: The Nernst potential should not change with stirring!)

Let us now add a small amount of Cu^{2+} ions. They adsorb at the Zn surface and catalyze the reduction reaction (6.26). The i_c will increase (Fig. 6.8b), resulting in the shift of the E_{mix} in the positive direction. (Note: Nernst potential should not shift with adsorption or with change of the heterogeneous rate constant for the interfacial process!)

Therefore, changes of "zero-current" potential with stirring, adsorption, or changes in the electrode surface are unmistakable signs of mixed potential and can serve as the first warning of experimental artifact.
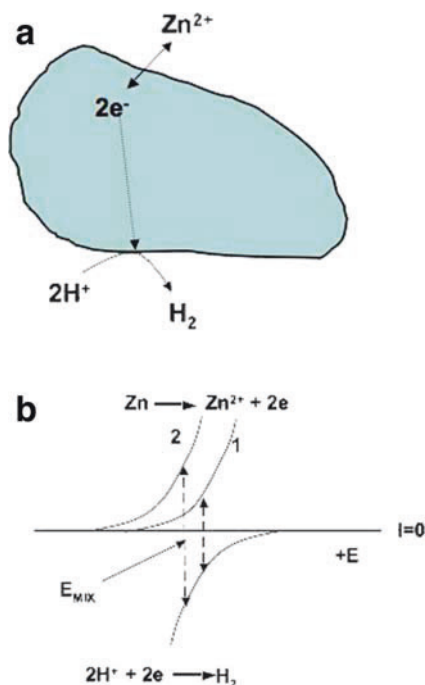
Fig. 6.8 (**a**) Mixed potential formed on a lump of Zn dissolving in acid and (**b**) shift of the mixed potential in negative direction upon addition of Cu^{2+} ions (curve 2) catalyzing the reduction reaction

6.2.2 Components of Potentiometric Cells

The need for a rather lengthy discussion of the liquid junction potential was prompted by the fact that, in potentiometry, the information is obtained from the measurement of the cell voltage.

$$E_{cell} = \pi_{ind} - (\pi_{ref} \pm E_j) \tag{6.27}$$

Fifty percent of this information comes from the indicator electrode and the other half from the reference system, including the liquid junction. The inescapable fact is that the electrical circuit has to be completed and that the probability of failure is equally distributed between the three terms in (6.27).

6.2.2.1 Reference Electrodes

There are three requirements that a good reference potential must satisfy. It must be stable, reversible, and reproducible. In this context, *stable* means that it will not change when the composition of the sample changes. *Reversible* means that it will

return rapidly to its equilibrium value after a small transient perturbation. This condition implies a low charge-transfer resistance R_{ct}. The *reproducibility* requirement also means that the same electrode potential will always be obtained when the reference electrode is constructed from the same electrode/solution combinations. The requirement of reversibility is easily satisfied by choosing electrochemical reactions that are very fast, that is, have a high exchange current density ($>10^{-3}$ A cm^{-2}) and the species involved in the charge-transfer process are chemically stable.

There are three types of reference electrodes discussed: reference electrodes of the first kind, reference electrodes of the second kind, and redox reference electrodes. The first two are used with potentiometric chemical sensors, whereas the last one helps us to get around the difficult problem of comparing potentials in different solvents. There is also a pseudo-reference electrode that does not have a stable, defined, reproducible potential. It serves only as the signal return to satisfy the condition of closing the electrical circuit (see Section 5.2). Because the liquid junction always causes some leakage of the internal solution, electrodes of the first kind are particularly affected.

First Kind

The potential of an electrode of the first kind is determined by the redox equilibrium between the cation with its metal. The simplest example is the equilibrium

$$Ag^+ + e^- = Ag^0 \tag{6.27}$$

In terms of (6.6), the Nernst potential is

$$\pi = \pi^0_{Ag/Ag^+} + \frac{2.303 \Re T}{F} \log a_{Ag^+} \pi^{\tilde{0}} = +0.799 \tag{6.28}$$

A reference electrode based on (6.28) can be realized by immersing Ag wire in a solution of $AgNO_3$. In order to maintain its activity constant this solution is placed in the inner reference electrode compartment, which is then connected to the sample solution through the liquid junction. Because of its good solubility, silver nitrate has often been used in reference electrodes for nonaqueous solvents.

In order to avoid confusion in reporting of electrochemical data, the Standard Hydrogen Electrode (SHE) has been selected as the primary reference electrode, which means that its standard potential has been arbitrarily set to zero, thus establishing the hydrogen scale of standard potentials. It is based on the equilibrium

$$H^+ + e \Leftrightarrow \frac{1}{2} H_2 \tag{6.29}$$

Its realization is shown in Fig. 6.9. This electrode is impractical and secondary reference electrodes such as those discussed in these sections are used instead.
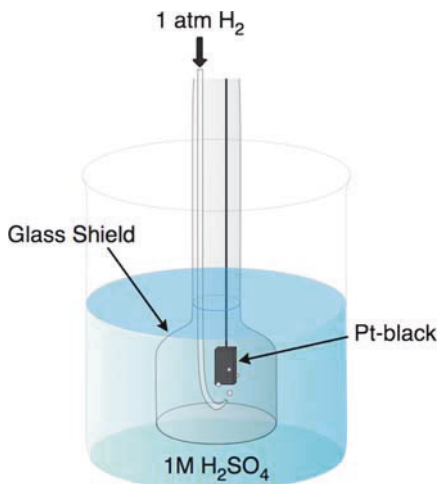
Fig. 6.9 Schematic of the standard hydrogen electrode

Second Kind

These electrodes are based on two equilibria: the electrochemical equilibrium involving formation of the interfacial potential and the solubility equilibrium between the cation and its sparsely soluble salt. The most popular electrode of this type is the silver/silver chloride electrode. The electrochemical equilibrium is the same as for the Ag/Ag^+ electrode described above (6.27) and the solubility equilibrium is

$$AgCl \overset{K_{sp}}{\Leftrightarrow} Ag^+ + Cl^- \tag{6.30}$$

Because AgCl is relatively insoluble, it maintains the constant activity of silver ions at the electrode.

$$AgCl \overset{K_{sp}}{\Leftrightarrow} Ag^+ + Cl^- \tag{6.31}$$

At 25°C the solubility product is $K_{sp} = 1.8 \times 10^{-10}$. It is defined as

$$K_{sp} = a_{Ag^+} a_{Cl^-} \tag{6.32}$$

Combining (6.28) and (6.32) yields the potential for the silver/silver chloride electrode.

$$\pi = \pi^0_{Ag/Ag^+} + \frac{2.303 \Re T}{F} \log K_{sp} - \frac{2.303 \Re T}{F} \log a_{Cl^-} \tag{6.33}$$

Note that the first two terms on the right-hand side of (6.33) combine to

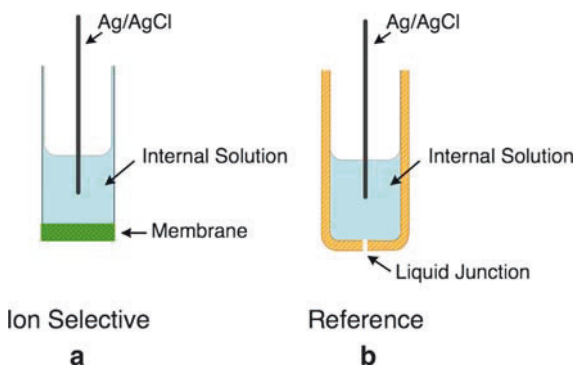$$\pi^0_{Ag/AgCl} = \pi^0_{Ag/Ag^+} + \frac{2.303 \Re T}{F} \log K_{sp}. \tag{6.34}$$

Fig. 6.10 (**a**) Classical arrangement of ISE in which Ag/AgCl electrode is used as internal reference. (**b**) Ag/AgCl electrode with liquid junction for use as external reference electrode

Table 6.1 Potentials of Ag/AgCl reference electrode against NHE

Concentration of KCl (M)	Potential vs NHE
0.01	+0.343
0.1	+0.288
1.0	+0.235
Saturated @25°C	+0.199

which, at 25°C, has value +0.223 V against the hydrogen electrode. The potential of this electrode is a function of the activity of chloride ions, which must be kept constant in order to satisfy the requirement of stability. Thus, the Ag/AgCl electrodes must be again kept in a separate compartment of defined and constant activity of chloride ion. The schematic diagram of a silver/silver chloride electrode, which is used as an internal reference (Fig. 6.10a), and an external reference electrode is shown in Fig. 6.10b.

The potential of the reference electrode clearly depends on the concentration of the internal solution. It is preferable to select the composition of the internal solution such that it matches the osmolality of the sample solution as much as possible in order to prevent flux of solvent through the junction. Values of reference electrode potentials for the Ag/AgCl electrode with different filling solutions are shown in Table 6.1.

Another popular electrode of the second kind, common in the older literature, is the calomel electrode. It is again based on redox and solubility equilibria.

$$Hg^+ + e \Leftrightarrow Hg^0 \tag{6.35}$$

and

$$2Hg^+ + 2Cl^- \overset{K_{sp}}{\Leftrightarrow} Hg_2Cl_2 \tag{6.36}$$
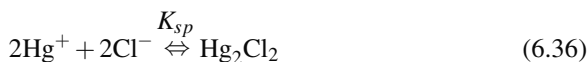
The solubility product of the Hg_2Cl_2 (calomel) is very low ($K_{sp} = 1.3 \times 10^{-17}$). The potential of this electrode is again determined by the concentration of the chloride ion in the inner compartment. When a saturated solution of KCl is used its potential against the SHE is $\pi = +241\,mV$. Use of a saturated KCl solution hides a certain danger: the higher temperature sensitivity, which is due to the temperature effect on solubility.

It is customary to write the electrochemical cell with interfaces shown as "/" and junctions as "//". Thus, for a simple cell consisting of a reference electrode and an indicator electrode, we write

$$Ag/AgCl/(c)KCl//sample/indicator\ electrode$$

If another compartment filled with, for example, $NaNO_3$, is used in order to prevent contamination of the sample with KCl, the double-junction electrode is written as

$$Ag/AgCl/(c)KCl//(c)NaNO_3//sample/indicator\ electrode$$
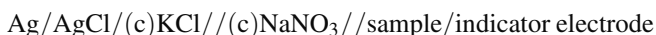
Redox Reference Electrode

This approach is used as a stepping stone when it is necessary to relate the potential obtained in a nonaqueous solution to an (aqueous) standard hydrogen electrode scale. It is known as the "internal redox reference" approach. The potential of an electrochemically well-behaved redox couple, which is fast and stable (e.g., ferrocene/ferricenium) is obtained by some voltammetric (see Chapter 8) method at an inert (e.g., Pt) working electrode and in the solvent and electrolyte of interest. The voltammogram is recorded against a reference electrode of the first kind (REF_{AUX}), which has stable, but otherwise "unknown," potential in that solvent. Typically, silver wire is used for that purpose. It is assumed (Strehlow and Wendt, 1961) that this reference redox couple is not significantly affected by the solvation. In other words, its standard potential is the same on the standard hydrogen scale (aqueous) as it is in any other solvent. The same REF_{AUX} is thus "calibrated" against this redox couple and can be used in the same solvent for the actual potentiometric measurement. Potentials measured against such an electrode are commonly reported as "... against Fe/Fec^+ redox couple." Obviously, no liquid junctions are involved in this approach.

Another form of redox reference electrode is similar to the electrode of the first kind. In this case the inert metal (e.g., Pt, Au, or C) is used as the inner electrode and a stable and soluble redox couple is placed inside the inner reference electrode compartment. A normal liquid junction is used in this type of reference electrode. Unlike the electrode of the first kind, the redox reference electrode is relatively immune to changes in concentration inside the reference electrode compartment because it is the ratio of the reduced/oxidized form of redox couple that determines the potential and not the absolute concentrations. However, redox reference electrodes are sensitive to changes of concentration of oxygen and other redox species.

Miniaturization of Reference Electrodes

Obviously, the volume of the inner compartment plays a critical role in the longevity
of the reference electrode and in its miniaturization. Because the liquid junction is
an open channel (i.e., electrical junction) through which the ions can communicate
between the compartment and the sample, it causes the change of activity and of the
potential. How rapidly this change takes place depends upon the volume of the inner
compartment, on the construction of the junction, and on the difference of the chem-
ical composition of the sample and the reference compartment solution. Generally
speaking, smaller volume, low junction resistance, and large chemical difference
between the sample and the reference solution result in a shorter usable lifetime. The
volume of the solution compartment in a practical laboratory reference electrode is
usually between 2–5 mL. Because the liquid junction is an open channel, there is
an outflow of the internal reference solution of approximately $2\,nL\,h^{-1}$ (Fig. 6.11).
This flow helps to stabilize the liquid junction potential (Dohner et al., 1986). The
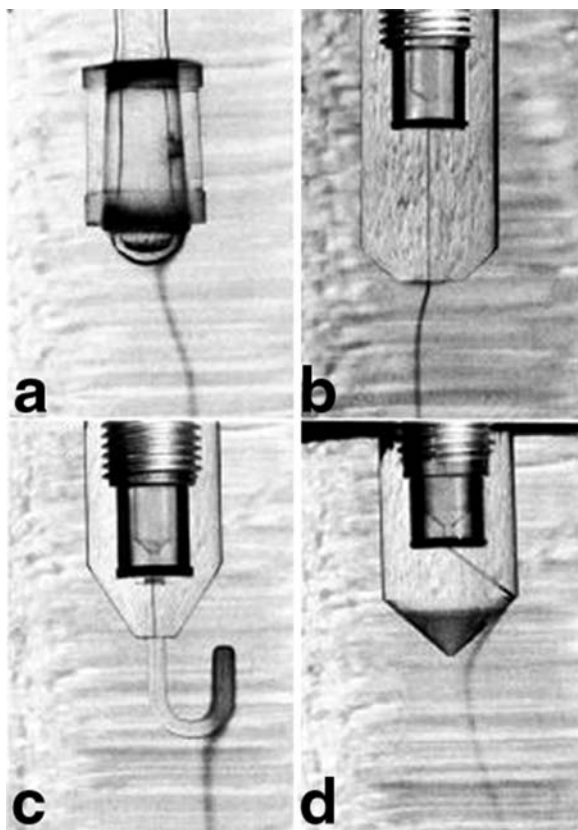
Fig. 6.11 Outflow of internal filling solution from various types of liquid junctions stabilizes the
E_j, but contaminates the sample

lifetime of such a reference electrode depends on the conditions under which it is used, but it often exceeds 1 month of continuous use.

Volume has been the most significant limitation on the size and construction of microreference electrodes, a limitation that complements the small size of the microfabricated ion sensors (Section 6.2.3.2). There have been many attempts to prepare a "liquid junction free" microreference electrode that would be comparable in size with the integrated ion sensors, such as ion-sensitive field-effect transistors (Section 6.2.3.2). These attempts have followed broadly three lines of reasoning: scaling down of a macroscopic reference electrode (Comte and Janata, 1978; Smith and Scott, 1986), elimination of the reference solution compartment while preserving the internal element structure (e.g., Ag/AgCl), and utilization of "inert" materials such as polyfluorinated hydrocarbons and the like, particularly in the so-called "reference FET" configuration.

The last type clearly violates the requirement of high exchange current density, which leads to instability of the potential. There is always a potential at any interface. Unless that potential is dominated by a high exchange current density, it cannot be stable. The second approach (internal element only) violates the condition of the constant activity of the reference ion (e.g., a_{Cl}). Although such an electrode will form a good ohmic contact with the solution, its potential will change with the composition of the solution. It can serve only as a pseudo-reference electrode. The first type, the miniature conventional reference electrode, can function but only for a short period of time, depending on the actual volume of the reference solution compartment and the conditions under which it is used. The usable lifetime is only minutes to hours. Such electrodes cannot sustain an outflow of the internal solution, which makes the liquid junction potential unstable in practical applications. In short, there are no true reference electrodes available at the micro scale.

However, an interesting approach has been recently reported that involves a highly hydrophobic membrane containing room-temperature ionic liquids (Maminska et al., 2006). This membrane has good electrical conductivity, long-term stability, and responds only to the most hydrophobic anions. Thus, although still a pseudo-reference electrode, it behaves well in solutions not containing hydrophobic anions. It serves as an illustration of how judicious design of the electrode phase can mitigate one of the most persistent problems of potentiometry. It can be said that there have been many pseudo-reference electrodes published and claimed to be "liquid junction free," but upon closer examination their usually unpublicized limitations become apparent. It is the result of the fact that a compromise always has to be made between "nonselectivity" of the junction (i.e., approximately equal and low charge-transfer resistance of all exchanging ions and electrons) and the volume of the liquid junction compartment.

It is often necessary to compare the potentials obtained with one reference electrode to those obtained with another reference electrode. It is important that the electrodes in question use the same solvent and that the contribution from the liquid junction potential is minimized. Otherwise, we have to use the internal reference redox couple. The comparison can be done according to the diagram shown in
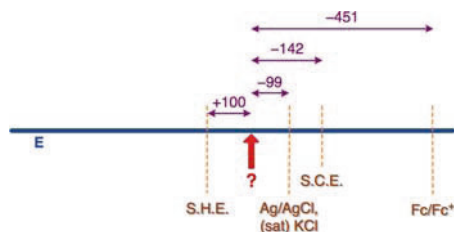
Fig. 6.12 Conversion of an unknown (?) potential measured with one reference electrode to another reference electrode scale (all values are in millivolts)

Fig. 6.12, in which potential E_x has been measured against a secondary reference electrode and it is necessary to convert it to the standard hydrogen electrode scale which defines the "zero". Thus, if the potential measured against Ag/AgCl (sat.) KCl//... is -99 mV, it is -142 mV versus SCE, -451 mV versus Fc/Fc$^+$ redox electrode, but $+100$ mV versus SHE.

6.2.2.2 Ion-Selective Electrodes

The ion-selective membrane is the key component of all potentiometric ion sensors. It establishes the preference with which the sensor responds to the ion of interest in the presence of various other ionic components of the sample. By definition, the ion-selective membrane forms a nonpolarized interface with the solution. If the interface is permeable to only one ion, the potential difference at that interface is expressed by the Nernst equation (6.6). If more than one ion can permeate, the interface can be anything between the liquid junction and the mixed potential. The key distinguishing feature is the absolute magnitude of the total exchange current density.

On the other hand, the ion selectivity depends on the relative magnitude of the exchange current density of the analyte ion x relative to the exchange current densities of the interferant ions j. The higher this ratio is, the more selective the membrane. A well-behaved membrane (i.e., the one which is stable, reproducible, immune to adsorption and stirring effects, and is also selective) has both high absolute exchange current density and high relative exchange current density. In contrast, a poorly behaving membrane with low exchange current density may still show some selectivity, but it will be affected by all the above factors. Although the formulation of the selectivity for ion sensors in terms of exchange current densities is conceptually correct, so far it has not led to the explicit expression for, or the explanation of, the physical meaning of the selectivity constant $K_{x,j}^{\text{pot}}$.

$$E_{\text{out}} = \frac{2.3\Re T}{z_x F} \log \left(a_x + \sum_{j i} K_{x,j}^{\text{pot}} a_j^{z_x/z_j} \right) \tag{6.37}$$

It is the same as the equation of general response formulated earlier (1.17), except that the potentiometric transfer function $\Re$ is now given by the Nernst equation. It

can be seen from (6.36) that the membrane is more selective for the determinand x the lower the value of each selectivity coefficient $K_{x,i}$. Thus, our ability to optimize is tantamount to designing better and more selective membranes. This depends on the identification of the mechanistic details of the ion partitioning processes, which affect the selectivity constant. A detailed treatment of this subject can be found in specialized books (Morf, 1981; Koryta and Stulik, 1983; Buck, 1981).

Ion-selective membranes are usually classified according to the nature of their physical state as solid or liquid membranes. The binding sites can be either ionic fixed binding sites, mobile ion-exchange sites, or neutral ionophores. The binding sites are incorporated in the membrane matrix, which determines the internal dielectric constant, lipophilicity, transport, and other mechanical and electrical properties. Although ion-selective membranes are usually discussed in the context of sensing in an aqueous environment, water is not the only usable medium in which these sensors can operate. There have been many successful applications of ISE in non-aqueous media and at high temperatures and pressures. The suitability of any given sensor for such applications is usually dictated by its construction and by the material aspects, rather than by some fundamental restrictions on the operation principles of the device.

6.2.2.2.1 Solid Membrane Electrodes

Glass Electrode

The oldest and the most important member of this family is the glass electrode. The proton-binding anionic sites are created by the defects in the SiO_2 matrix, such as the cationic vacancies due to the nonsilicon constituents of the glass (Eisenman, 1966; Fig. 6.13). For example, in the Corning 015 glass the bulk matrix is silica (72%) with 22% Na_2O and 6% CaO added.
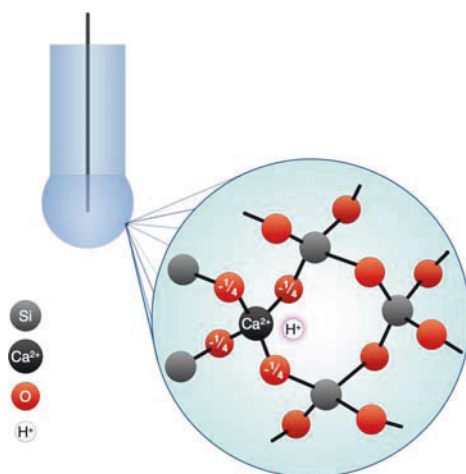

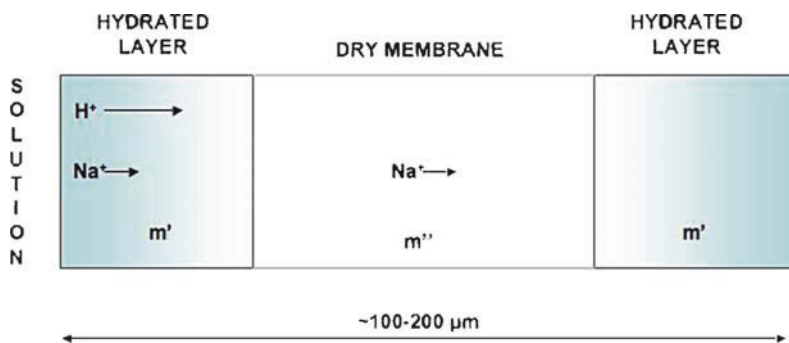**Fig. 6.13** Defects in silica matrix of glass electrode act as binding sites for protons

Fig. 6.14 Glass membrane with hydrated layers (m′) on both sides of the dry interior (m″)

Tens of different specialized glass compositions have been published and patented. The main goals are improvement of selectivity and lower overall resistance.

When the glass membrane is exposed to water, a hydrated layer, approximately 50–100 nm thick, is formed at its interface. In addition to water, the chemical composition of the glass in this layer is the same as that in dry bulk. The concentration of the anionic binding sites is estimated between 3 and 10 M. The membrane is usually blown into a bulb of a typical thickness of the wall 50–200 μm. The optimum thickness of the wall is a compromise between mechanical stability and the electrical resistance. The latter is typically on the order of 10 MΩ. The interior of this bulb is sealed and contains the internal reference electrode. Thus, the glass membrane is bathed on both sides by solution and a similar hydrated layer develops on the inside of the glass bulb as well (Fig. 6.14).

This is called a symmetrical membrane arrangement. It is the original version found in all early ion-selective electrodes. The cell notation can be written as

$$Ag/AgCl, 0.1M\ HCl/hydrated\ layer/dry\ glass/hydrated\ layer/sample\ldots$$
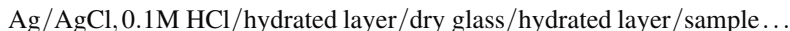
The symmetrical arrangement cannot be easily scaled down, but it has one major advantage: most nonidealities cancel out. It is a good starting point for discussion of all ion potentiometric sensors.

There are two processes that take place during the interaction of the hydrated layer and the sample: the ion exchange and the diffusion of all participating ions. We show that they both contribute to the value of the membrane potential and of the selectivity coefficient. Let us consider a simple case of only two univalent ions Na^+ and H^+ exchanging between the solution (S) and the hydrated layer (m′) of the membrane, according to the equation:

$$H^+(m') + Na^+(S) \overset{K_{ex}}{\Leftrightarrow} H^+(S) + Na^+(m') \tag{6.38}$$
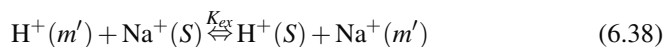
Therefore, their electrochemical potentials in these two phases must be equal:

$$\tilde{\mu}_{H+}^{S} = \tilde{\mu}_{H+}^{m'} \tag{6.39a}$$

and

$$\tilde{\mu}_{\text{Na}^+}^S = \tilde{\mu}_{\text{Na}^+}^{m'} \tag{6.39b}$$

The equilibrium ion-exchange constant K_{ex} for this process is

$$K_{\text{ex}} = \frac{a_{\text{Na}^+}^{m'} a_{\text{H}^+}^S}{a_{\text{H}^+}^{m'} a_{\text{Na}^+}^S} \tag{6.40}$$

The electrochemical potentials can be expanded and rearranged in the usual way (6.3), to yield the expression for the interfacial potential.

$$\pi = \varphi^m - \varphi^S = \frac{\mu_{\text{H}^+}^{0,S} - \mu_{\text{H}^+}^{0,m'}}{F} + \frac{\Re T}{F} \ln \frac{a_{\text{H}^+}^S}{a_{\text{H}^+}^{m'}} \tag{6.41}$$

Because the standard potentials for the proton in the solution and in the hydrated layer are the same ($\mu_{\text{H}^+}^{0,S} = \mu_{\text{H}^+}^{0,m'}$), the first term on the right-hand side of (6.39) cancels out.

The hydrated layer has finite thickness, therefore the exchanging ions can diffuse inside this layer, although their mobility is quite low compared to that in water ($u \sim 10^{-11}\,\text{cm}^2\,\text{s}^{-1}\,\text{V}^{-1}$). As we have seen in the liquid junction, diffusion of ions with different velocities results in charge separation and formation of the potential. In this case, the potential is called the diffusion potential and it is synonymous with the junction potential discussed earlier. It can be described by the equation developed for the linear diffusion gradient, that is, by the Henderson equation (6.24). Because we are dealing with uni-univalent electrolytes, the multiplier cancels out and this diffusion potential can be written as

$$\pi_{\text{diff}} = \frac{\Re T}{F} \ln \frac{u_{\text{H}^+} a_{\text{H}^+}^{m'} + u_{\text{Na}^+} a_{\text{Na}^+}^{m'}}{u_{\text{Na}^+} a_{\text{H}^+}^{m''}} \tag{6.42}$$

The charge transport in the dry glass (m'') is by the mobile sodium ions, because protons have near zero mobility in dry glass. The symmetry of the membrane arrangement means that the same processes apply at the interior interface of the glass membrane.

In reality, the cell voltage comprises all potential differences between the interior of the solution and the interior reference electrode. Therefore, all ion exchange and diffusion potentials have to be added to yield the membrane potential of the glass electrode.

$$\pi_{\text{m}} = \frac{\Re T}{F} \ln \frac{(u_{\text{Na}^+}/u_{\text{H}^+}) K_{\text{H}^+,\text{Na}^+} a_{\text{Na}^+}^S + a_{\text{H}^+}^S}{(u_{\text{Na}^+}/u_{\text{H}^+}) K_{\text{H}^+,\text{Na}^+} a_{\text{Na}^+}^{\text{int}} + a_{\text{H}^+}^{\text{int}}} \tag{6.43}$$

Because the interior reference electrode is sealed, all terms in the denominator of (6.43) of this semiempirical equation are constant. It is called the Nikolskij and Eisenman equation (Morf, 1981).

$$\pi_{\mathrm{m}} = \mathrm{const.} + \frac{\Re T}{F} \ln \left(a_{\mathrm{H}^+}^S + K_{\mathrm{Na}^+,\mathrm{H}^+}^{\mathrm{pot}} a_{\mathrm{Na}^+}^S \right) \tag{6.44}$$

The potentiometric selectivity coefficient $K_{\mathrm{Na}^+,\mathrm{H}^+}^{\mathrm{pot}}$ is defined as

$$K_{\mathrm{Na}^+,\mathrm{H}^+}^{\mathrm{pot}} = (u_{\mathrm{Na}^+}/u_{\mathrm{H}^+})K_{\mathrm{ex}} \tag{6.45}$$

It expresses the preference of the glass electrode for the hydrogen ion with respect to the sodium ion.

The glass electrode is the most selective electrode known and it has the widest dynamic range of all chemical sensors, experimentally estimated at over 36 decades of hydrogen ion activity! The reason for this astonishing performance lies in the multiplicity of proton-binding sites in the hydrated layer of the glass (Chapter 1). Because sodium ion and alkali metal ions in general are involved in the ion-exchange equilibrium (6.37), they present the biggest selectivity challenge. The loss of selectivity in solutions of high activity of alkali metal ions is known as the "alkaline error" of the glass electrode. Sensitivity to sodium ions can be mitigated by changing the composition of the glass, that is, by replacing sodium oxide with lithium oxide. However, that increases its sensitivity for Li^+. It is a never-ending problem, which always comes down to relative values of exchange current densities for the species involved in the charge-transfer process.

The importance of the Nikolskij–Eisenman equation is much more than just historical. It shows that two processes are involved in the origin of the selectivity coefficient, the ion-exchange equilibrium (6.38), and the mobilities of the ions. Therefore, it points the way for optimization of the selectivity coefficient, through the optimization of the mechanisms that give rise to it. It also highlights the conceptual connection between the mixed potential and the true equilibrium potential. If the exchange current density for one ion is very high relative to the other partial exchange current densities, the E_{m} lies close to the equilibrium potential of this dominating charge-transfer process and the membrane behaves as a true equilibrium sensor. Despite the fact that the complicated algebra prevents the potentiometric selectivity coefficients being expressed in terms of the partial exchange current densities, it is an attractive concept because it allows at least correct interpretation of potentiometric behavior of low exchange current density membrane electrodes. Therefore, the exchange current densities and the selectivity coefficients are closely related, as is evident from the data experimentally obtained for a K^+ selective membrane (Table 6.2).

The "const" in (6.44) points to the importance of the construction aspects of ion sensors. Even if the glass membrane were placed between two identical solutions, the E_{m} would not be zero. This is due to the fact that the membrane develops an "asymmetry potential," which arises from the different degrees of mechanical stress at the interior and exterior interfaces of the glass. This affects the exchange current densities. We return later to this point, in the discussion of ion sensors with asymmetric membrane.

Table 6.2 Estimates of Apparent Standard Exchange Current Densities $i_0^{0,a}$ Exchange Current Ratios i_0, $i/_0$,K and Selectivity Coefficients Obtained by Different Methods[b]

Solution	i_0^0 A/cm^2	0.1 M solutions		10^{-4} M solutions	
		$i_0, I/i_0, K$	K_{K-I}	$i_0, I/i_0,$K	K_{K-I}
KCl	2.1×10^{-3}	1	1	1	1
RbCl	4.7×10^{-3}	3.2	2.4	0.6	0.5
			2.5		0.4
CsCl	5.3×10^{-4}	0.32	0.34	0.5	0.5
			0.32		0.4
NH$_4$Cl	4.2×10^{-5}	4×10^{-2}	2×10^{-2}	0.6	0.1
			1.4×10^{-2}		0.08
			10^{-26}		
NaCl	4.2×10^{-6}	10^{-3}	2×10^{-4}	0.5	5×10^{-2}
			4×10^{-5}		3×10^{-2}
			2.5×10^{-5c}		
LiCl	2.6×10^{-6}	7×10^{-4}	5.3×10^{-5}	0.4	3.4×10^{-2}
			2×10^{-3}		2×10^{-2}
N(CH$_3$)$_4$Cl	2.1×10^{-6}	6.6×10^{-4}	4.6×10^{-6}	0.3	3×10^{-2}
			1.7×10^{-5}		2×10^{-2}
N(C$_2$H$_5$)$_4$Cl	1.2×10^{-6}	2.5×10^{-4}	6.7×10^{-5}	0.2	3×10^{-2}
			3.1×10^{-6}		2×10^{-2}

[a]Membrane phase, 2.7×10^{-3} M vallnomycin in diphenyl ether; geometrical surface, $-0.012 \, \text{cm}^3$
[b]First value, separate solution of the same concentration; second value, separate solution with the same potential; third value, mixed solution method
[c]Constant concentration of the interfering ion, $I = 0.1 \, \text{M}$

From (6.44), we see that in order to obtain a selective membrane, the value of the ion-exchange constant must be small and the sodium ion mobility in the hydrated layer relative to that of the hydrogen ion must also be small. The expansion of the selectivity coefficient to include selectivity to other ions involves inclusion of more complex ion-exchange equilibria, and the use of a more complex form of the Nernst–Planck equation. This rapidly leads to intractable algebra that requires numerical solution (Franceschetti et al., 1991; Kucza et al., 2006). Nevertheless, the concept of the physical origin of the selectivity coefficient remains the same. Electrochemical impedance spectroscopy has been successfully used in analysis of the ISE function (Gabrielli et al., 2004).

Fluoride Electrode

The glass electrode is so important that it deserves a special section in this book. However, there are other solid-state ISEs that have excellent performance, and one of them is the fluoride-selective electrode. There are only few analytical methods that allow simple and selective determination of fluoride ion. For this reason, the ion-selective electrode is one of the most important analytical tools. Although it was

introduced much later than the glass electrode, it is generally regarded as the first modern ion-selective electrode (Frant and Ross, Jr., 1966). It began the renaissance of zero-current potentiometry which continues to the present time. It consists of a single crystal of lanthanum fluoride doped with europium ions and has the general form as shown in Fig. 6.10a. The membrane is based on the following equilibrium.

$$LaF_3 = La^{3+} + 3F^-$$ (6.46)

Its potential is given by (6.45).

$$\pi = \pi^0_{F^-/La_3F} - \frac{\Re T}{F} \ln \left(a_{F^-} + K^{\text{pot}}_{F^-,OH^-} a_{OH^-} \right)$$ (6.47)

Here, the potentiometric selectivity coefficient is given with respect to the hydroxyl ion. Single-crystal lanthanum fluoride is a wide bandgap semiconductor in which the electrical conductivity is due only to the hopping mobility of fluoride ions through the defects in the crystal. It does not respond to the La^{3+} ion because of the slow ion exchange of that ion. Hydroxyl ion is the only other ion that has appreciable mobility, and is the only known interference. For this reason, the measurements with a fluoride electrode are always done below pH 7, which circumvents this interference. As shown later, the consideration of ionic and/or electronic conductivity of the membrane plays a critical role also in the design of the internal contact in nonsymmetric potentiometric sensors.

Other Solid-State Ion-Selective Electrodes

Other types of solid-state membranes include single crystals of sparsely soluble salts and are often called heterogeneous membranes, in which the insoluble salt is embedded in some inert polymer matrix. Obviously, in order for these membranes to be at equilibrium they should be in a saturated solution. In practice, these membranes are used in solutions that are below saturation. In that case, the "insoluble salt" slowly dissolves.

The key common property of such membranes is their ionic conductivity. An example is the Ag/AgCl electrode discussed previously. Although the primary charge-transfer reaction at this interface is that of the silver ion (6.27), the electrode will respond also to chloride ion (6.33), because of the low solubility product of AgCl. Let us now consider what happens if other ions that also form an insoluble silver salt are present in the solution. If the solubility product of the other salt (AgX) is lower than that of silver chloride and/or the activity a_x is sufficiently high

$$\frac{K_{\text{sp(AgCl)}}}{K_{\text{sp(AgX)}}} > \frac{a_{Cl^-}}{a_{X^-}}$$ (6.48)

The membrane undergoes metathesis and its surface becomes gradually partially covered with the solid phase of AgX. Clearly, this is a nonequilibrium situation that leads to formation of mixed potential. It also underscores the nonequilibrium

Table 6.3 Membranes for Solid-State Ion-Selective Electrodes

Primary ion	Orion electrode[a]	Homogeneous membranes[b]	Heterogeneous membranes[c]
F^-	LaF_3		–
Cl^-	$AgCl/Ag_2S$	$Hg_2Cl_2/HgS, AgCl$	$AgCl$
Br^-	$AgBr/Ag_2S$	$Hg_2Br_2/HgS, AgBr$	$AgBr$
I^-			
CN^-	AgI/Ag_2S	AgI	AgI
Hg^{2+}			
SCN^-	$AgSCN/Ag_2S$	$Hg_2(SCN)_2/HgS, AgSCN$	$AgSCN$
S^{2-}	Ag_2S	Ag_2S	Ag_2S
Cu^{2+}	Cu_xS/Ag_2S	$CuSe$	$Cu_xS, Cu_xS/Ag_2S$
Pb^{2+}	PbS/Ag_2S		PbS/Ag_2S
Cd^{2+}	CdS/Ag_2S		CdS/Ag_2S

[a]Typical commercial electrode
[b]Pressed pellet or a single crystal
[c]Solid powder in an organic matrix

nature of the Nikolskij–Eisenman equation. For membranes based on silver salts, this type of interference is particularly severe if the sulfide ion is present because of the exceptionally low solubility product of Ag_2S. On the other hand, the low value of the solubility product of silver sulfide can be used to advantage. It is possible to prepare membranes in which Ag_2S or HgS are used as a matrix and a more soluble salt is present as an additive. Examples of such membranes are given in Table 6.3.

Let us consider mixed PbS–Ag_2S membrane, for which the respective solubility products at 25°C are

$$K_{sp(Ag_2S)} = a_{Ag^+}^2 a_{S^{2-}} = 6 \times 10^{-50} \tag{6.49a}$$

and

$$K_{sp(PbS)} = a_{Pb^{2+}}^2 a_{S^{2-}} = 2.5 \times 10^{-27} \tag{6.49b}$$

The activity of lead ion is

$$a_{Pb^{2+}} = \frac{K_{sp(PbS)}}{K_{sp(Ag_2S)}} a_{Ag}^2 \tag{6.50}$$

This activity of lead would be present in the saturated equilibrium solution, due to the dissolution of the membrane. Therefore, it represents the detection limit. Because silver ion is the charge-transferring species, the membrane potential must be written in terms of the silver ion activity (6.37), which is obtained from (6.49). Thus

$$\pi_m = \pi_{Ag/Ag^+}^0 + \frac{\Re T}{2F} \ln \frac{K_{sp(Ag_2S)}}{K_{sp(PbS)}} + \frac{\Re T}{2F} \ln \left(a_{Pb^{2+}} + K_{Ag^+,Pb^{2+}}^{pot} a_{Ag^+} \right) \tag{6.51}$$

The first two terms represent the "standard" potential for this electrode and $K^{\text{pot}}_{\text{Ag}^+,\text{Pb}^{2+}}$ is the potentiometric selectivity coefficient for Pb^{2+} with respect to Ag^+. Because Ag^+ is the only ion that can move in the Ag_2S matrix, the relative mobility is not a factor. The selectivity coefficient is then expressed as the ratio of the solubility products of the two salts

$$K_{\text{Pb/Ag}} = \frac{K_{\text{sp}}(\text{Ag}_2\text{S})}{K_{\text{sp}}(\text{PbS})} \tag{6.52}$$

The solubility of these silver salts is very low and the metathesis that takes place is also a very slow process. Nevertheless, (6.49) and (6.50) are not true equilibrium equations and have only semiquantitative meaning.

6.2.2.2.2 Liquid Membrane Electrodes

This title is historical and is slightly misleading because in the modern implementation these membranes are usually "solid". However, the binding "sites" are more or less mobile complexing agents called ionophores, which are dissolved in a suitable solvent and trapped in an organic matrix. Thus the membrane forms an organic gel. Because ion activity measurements are done predominantly in aqueous media the membranes are usually hydrophobic. Thus, the primary interaction between the ion in water and the hydrophobic membrane containing the ionophore is the extraction process, driven by the chemical affinity of the ion for the membrane. In principle, membranes of this type would function also in organic solvents and a few reports of such applications have been published. However, the major limitation is the solubility of the membrane itself and/or the leaching out of the membrane of the ionophore or of the plasticizer. In contrast, solid-state membranes based on inorganic materials can be used in organic solvents without limitations. The general arrangement of such an electrode again corresponds to Fig. 6.10a.

The most common polymer matrix is the pure, high-molecular weight poly(vinyl chloride). It constitutes typically 30% (by weight) of the membrane. Original liquid membrane ion-selective electrodes indeed contained liquid organic solvent with a suitable extracting agent dissolved in it. Because the liquid phase had to be retained, typically in some porous support, their construction was rather awkward. Introduction of PVC matrix and realization of the membrane in the form of doped organic gel was a significant turning point in the development of this type of potentiometric sensors (Moody and Thomas, 1978). The remaining major component of the membrane is the solvent (plasticizer), which ensures the mobility of the free and complexed ionophore, sets the dielectric constant, and provides suitable mechanical properties of the membrane. The ionophore usually comprises about 1% of the membrane, which corresponds to a concentration of binding sites of approximately 10^{-2} M. This is relatively low, as compared to the glass electrode. Other membrane components include large hydrophobic anions, such as tetrafluoroborate, and polysiloxane. They both extend the upper limit of the dynamic range by minimizing the Donnan exclusion failure (Iglehart et al., 1988). The

polysiloxane improves the adhesion of the membrane in asymmetric ion sensors (Li et al., 1988).

There are two kinds of ionophores: charged ones, which are called liquid ion exchangers, and neutral carriers. Because they are mobile in both the free and in the complexed form, mobilities of all species are again part of the selectivity coefficient together with the ion-exchange equilibrium constant. The best-known neutral ionophore is valinomycin (Fig. 6.15a) which shows a 1,000:1 selectivity for K^+ in preference to Na^+ and no pH dependence. In its uncomplexed form, it is electrically neutral. A better-known representative is di(n-octyl phenyl) phosphonate (Fig. 6.15c), which shows good selectivity for calcium ion and is relatively pH insensitive.

A large number of neutral ionopohores for cations and anions have been synthesized and many are commercially available (Morf and Simon, 1978). The organic solvent forming the gel membrane is hydrophobic and has a high dielectric constant. Examples of membrane constituents of liquid ISEs are given in Table 6.4.

An important consideration, particularly from the point of view of their use in solid-state sensors, is the optimum thickness of the membrane. For the membrane interior to be electrically neutral, the total thickness has to be greater than the combined thicknesses of the space charges that extend from both sides of the membrane
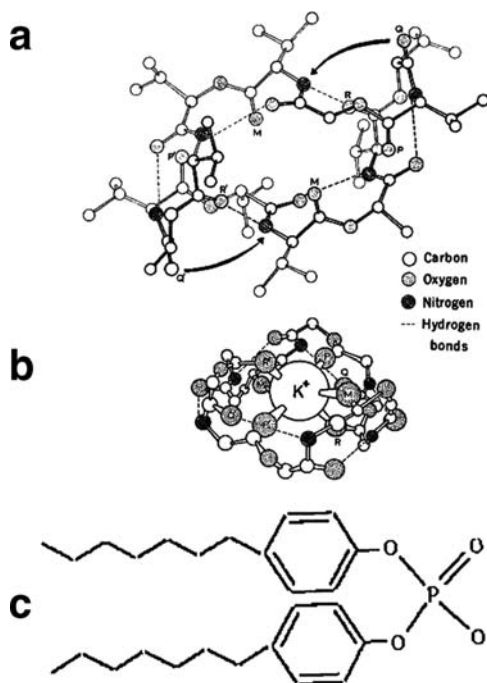
Fig. 6.15 (**a**, **b**) Neutral ionophore valinomycin with the cavity for binding potassium ion. (**c**) Charged ionophore di n-octyl phenyl phosphonate used in calcium ion-selective electrodes

Table 6.4 Electrodes based on liquid ion exchangers

Ion	Form of membrane	Active material	Solvent mediator
Ca^{2+}	Liquid	Calcium di-(n-decyl) phosphate	di-(n-octylphenyl) phosphonate
	Solid (PVC)	Calcium di(n-decyl) phosphate	i-(n-octylphenyl) phosphonate
NO_3^-	Solid (PVC)	Tris(substituted 1,10-phenan-throline) nickel(II) nitrate	p-Nitrocymene
ClO_4^-	Liquid	Tris(substituted 1,10-phenan-throline) iron(II) perchlorate	p-Nitrocymene
BF_4^-	Liquid	Tris(substituted 1,10-phenan-throline) nickel(II) tetrafluoroborate	p-Nitrocymene
Divalent cations (water hardness)	Liquid	Calcium di-(n-decyl) phosphate	Decanol
U(VI)	Solid(PVC)	di(2-ethylhexyl) phosphoric acid	Diamylamyl phosphonate
Cl^-	Liquid	Dimethyl-dioctadecylammonium chloride	

(Buck, 1981). The thickness of the space charge κ_Δ^{-1} is called the Debye length. It depends on temperature, on the dielectric constant ε of the membrane, and on the valency and concentration of the binding sites.

$$\kappa^{-1} = \left(\frac{\varepsilon \Re T}{z^2 F^2 \sum\limits_i C_i} \right)^{1/2} \tag{6.53}$$

It is typically on the order of several hundred nanometers. In practice the minimum thickness for polymeric membranes is $50\,\mu m$ or greater, which is far more than one would expect from (6.53). This is apparently due to the fact that these membranes hydrate in the bulk, thus increasing the dielectric constant. They also form a hydrated layer at the solution/membrane interface (Li et al., 1996) which affects their overall electrochemical properties and selectivities. Macroscopic ISEs use relatively thick membranes ($\sim 500\,\mu m$). In contrast, it is desirable to use thin membranes in the construction of asymmetric solid-state potentiometric ion sensors, in order to make their preparation compatible with the thin-layer preparation techniques.

6.2.3 Assembled Electrochemical Cell

Let us now return to the nonpolarized interface within the context of the working of the entire electrochemical cell, which we have to use in order to obtain useful information about concentration of fluoride ion, using (6.20). It is connected to a high-input impedance electrometer (e.g., $R > 10^{11}\Omega$), so that current cannot pass through it. This ensures that the condition of "zero current" is satisfied. The general schematic of ISE is shown in Fig. 6.10a. In the usual cell notation we can write for the complete cell

$$\text{Cu} \overset{\leftarrow}{/} \text{Ag} \underset{\text{Ag}^+}{\overset{\rightarrow}{/}} \text{AgCl} \underset{\text{Cl}^-}{\overset{\leftarrow}{/}} \text{KCl}, \text{KF}(c) \underset{\text{F}^-}{\overset{\leftarrow}{/}} \text{La}_3\text{F} \underset{\text{F}^-}{\overset{\leftarrow}{/}} \text{sample}(\uparrow x)\text{NaF} \underset{\text{Na}^+,\text{Cl}^-}{\overset{\leftrightarrow}{//}} \text{1MKCl} \underset{\text{Cl}^-}{\overset{\leftarrow}{/}} \text{AgCl} \underset{\text{Ag}^+}{\overset{\rightarrow}{/}} \text{Ag} \overset{\leftarrow}{/} \text{Cu}$$
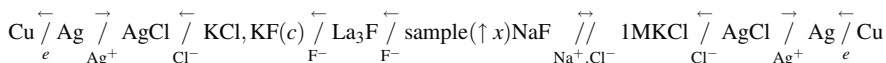
The arrows above and the symbols below the interfaces indicate the transfer of the charge at each interface when the concentration of NaF in the sample is abruptly increased. It is possible to estimate the actual number of ions that are required to establish the potential difference at the interfaces. A typical value for the double-layer capacitor is $10^{-5}\,\text{F}\,\text{cm}^{-2}$. If a potential difference of $\pi = 100\,\text{mV}$ is established at this interface, the double-layer capacitor must be charged by the charge $Q = \pi C_{\text{dl}} = 10^{-6}$ coulombs. From Faraday's law (6.3), we see that it corresponds to approximately $10^{-11}\,\text{mol}\,\text{cm}^{-2}$ or 10^{12} ions cm^{-2} of the electrode surface area. Thus, a finite amount of the potential determining ions is removed from the sample but this charge is replenished through the liquid junction, in order to maintain electroneutrality.

Although seemingly cumbersome, the process of identifying charge-transporting species across each interface in the cell is very important. This process ensures that the two fundamental conditions of potentiometric measurement mentioned in Chapter 5 are met. These are again: the condition of electroneutrality of each phase and the condition of a closed electrical circuit. The corollary of the second condition is the one-capacitor rule: *Only one capacitor in series is permitted in a potentiometric measurement*. That capacitor is the input capacitor of the voltmeter; all other interfaces must be resistive, that is, nonpolarized. In our example, upon increase of the fluoride ion concentration in the sample, charges are transferred across all interfaces and the cell voltage E_{cell} then appears across the input capacitor; the plate on the left is negatively charged, and the plate on the right is positively charged.

Let us obtain some additional information from this experiment by considering what would happen if we left out KF from the internal compartment. The AgCl/KCl interface would still be nonpolarized and chloride ions would transfer from the internal solution to the solid AgCl phase, but the $\text{KCl}/\text{La}_3\text{F}$ interface would be blocked (i.e., capacitive), because no charged species could transfer across it. Hence, the one-capacitor rule would be violated. The practical consequence would be uncontrollable drift of the E_{cell}.

The liquid junction is open to transfer of all ions. We have indicated that charge will be carried both by chloride anions from the right to the left and sodium cations from left to right. The potential of the junction remains constant.

As we make this concentration step, the potential profile through the reference solution compartment and through the liquid junction will be (approximately)

unchanged. The word "approximately" in the preceding sentence reflects the thermodynamic uncertainty that is associated with the liquid junction. As was pointed out previously, it is a nonequilibrium element because it is the site of the diffusion process. Nevertheless, within the limitation of this uncertainty it allows us to measure (approximately) the potential of the indicator electrode. Thus, in summary, use of the liquid junction gives us the practical possibility of measuring the potential of the indicator electrode at the cost of relaxation of thermodynamic rigor. Hence, the increase of the potential difference that appears at the La_3F/sample interface corresponds to the measurable change of the cell voltage, E_{cell}.

Ion-selective membranes can be used in two basic configurations. If the solution is placed on either side of the membrane, the arrangement (e.g., Fig. 6.16a) is *symmetrical*. It is found in conventional ion-selective electrodes in which the internal contact is realized by the solution in which the internal reference electrode is immersed. In the *nonsymmetrical* arrangement (Fig. 6.16b), one side of the membrane is contacted by the sample (usually aqueous), and the other side is interfaced with some solid material. Examples of this type are coated wire electrodes and Ion-Sensitive Field-Effect Transistors (ISFETs).
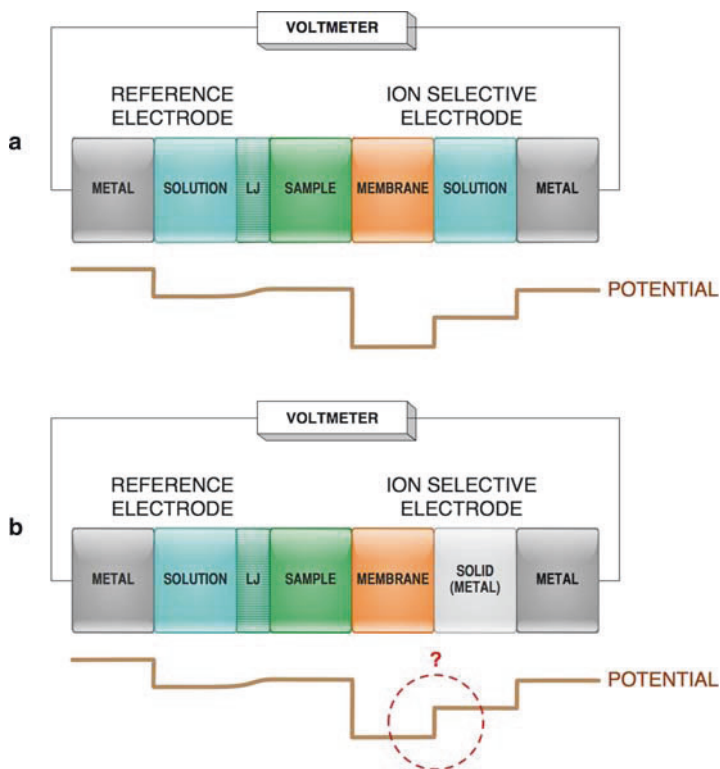
Fig. 6.16 (**a**) Symmetrical and (**b**) asymmetrical placement of ion-selective membrane. The interface in the circle, marked "?", requires particular attention

6.2.3.1 Symmetrical Ion-Selective Electrodes

The potential profile through the membrane that is placed between the sample and the internal reference solution was shown in Fig. 6.3. The composition of the internal solution can be optimized with respect to the membrane and the sample solution. In the interest of symmetry, it is advisable to use the same solvent inside the electrode as is in the sample. This solution also contains the analyte ion in the concentration, which is usually in the middle of the dynamic range of the response of the membrane. The ohmic contact with the internal reference electrode is provided by adding a salt that contains the appropriate ion that forms a fast reversible couple with the solid conductor. In recent designs, gel-forming polymers have been added into the internal compartment. They do not significantly alter the electrochemistry, but add mechanical stability and convenience of handling.

The obvious advantage of the symmetrical arrangement is that the processes at all internal interfaces can be well defined and that most nonidealities at the membrane/solution interface tend to cancel out. Because the volume of the internal reference compartment is typically a few milliliters, the electrode does not suffer from exposure to electrically neutral compounds that would penetrate the membrane and change the composition of this solution. This type of potentiometric ion sensor has been used in the majority of basic studies of ion-selective electrodes. Most commercial ion-selective electrodes are also of this type. The drawbacks of this arrangement are also related to the presence of the internal solution and to its volume. Mainly for this reason, it is not conveniently possible to miniaturize it and to integrate it into a multisensor package.

6.2.3.2 Asymmetric Membrane Ion Sensors

Symmetrical placement of the ion-selective membrane is typical for the conventional ISE. It helped us to define the operating principles of these sensors and most important, to highlight the importance of the interfaces. Although such electrodes are fundamentally sound and proven to be useful in practice, the future belongs to the miniaturized ion sensors. The reason for this is basic: there is neither surface area nor size restriction implied in the Nernst or in the Nikolskij–Eisenman equations. Moreover, multivariate analysis (Chapter 10) enhances the information content in chemical sensing. It is predicated by the miniaturization of individual sensors. The miniaturization has led to the development of potentiometric sensors with solid internal contact. They include Coated Wire Electrodes (CWE), hybrid ion sensors, and ion-sensitive field-effect transistors. The internal contact can be a conductor, semiconductor, or even an insulator. The price to be paid for the convenience of these sensors is in the more restrictive design parameters. These must be followed in order to obtain sensors with performance comparable to the conventional symmetrical ion-selective electrodes.

The two types of placement of ion-selective membrane are schematically shown in Fig. 6.16a and b. If designed correctly, they will perform the same sensing
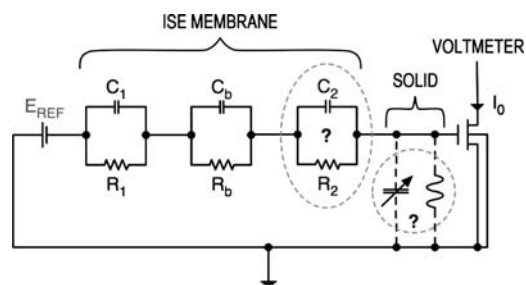
Fig. 6.17 Equivalent electrical circuit representing potentiometric sensor with asymmetrically placed membrane (as in Fig. 6.16b). The field-effect transistor input to the voltmeter is added

function. The key issue is again the interface between the membrane and the contact. The most convenient way to discuss this problem is to use the equivalent electrical circuit. Associated with the internal contact is the parasitic capacitance and resistance shown in the dashed circle in Figs. 6.16b and 6.18. In the analysis of these sensors, it is necessary to include the electrometer (or at least its input stage; Demoz et al., 1995). In most modern instruments, it is an Insulated Gate Field-Effect Transistor (IGFET) which has the input DC resistance of $10^{14}\,\Omega$ and the input capacitance on the order of a few picofarads.

Two limiting cases of the value of charge-transfer resistance are treated in Fig. 6.17. For a good ion-selective electrode, the charge-transfer resistance R_l is low ($i_0 > 10^{-3}$ A cm^{-2}; i.e., $R_{ct} < 25\,\Omega\,\mathrm{cm}^2$). This means that at least one charged species can transfer easily between the sample and the membrane and establish the reversible potential according to one of the mechanisms discussed for nonpolarized interfaces. The bulk membrane resistance R_b can be as high as $10^6\,\Omega\,\mathrm{gm}^2$. However, because no net current passes through the membrane the potential in the bulk of the membrane is uniform; that is, there is no electric field inside the membrane. For nontransient measurements, the bulk capacitor C_b can be neglected. In a conventional symmetrical ISE arrangement, the composition of the internal solution should be always chosen in such a way that the interfacial charge-transfer resistance R_2 is comparable to R_l. We have seen this in the discussion of the fluoride electrode above (Fig. 6.16a). Thus, a well-established potential profile exists throughout this structure. On the other hand, when the charge-transfer resistance at the internal interface (R_2) is very high, there is obviously no charge transfer between the internal conducting contact and the membrane. In this case, the input capacitor together with the interfacial capacitance C_2, together with the variable parasitic capacitance, form an undefined and variable capacitive divider. Now, the voltage that appears at the input capacitor of the electrometer depends, not only on the electrostatic potential of the membrane, but also on the undefined parasitic impedance associated with the internal contact. Thus, the presence of the interfacial capacitor C_2 violates the one-capacitor rule mentioned above. In a real situation the interfacial resistance R_2 will have some finite value and as the result such an electrode will drift.
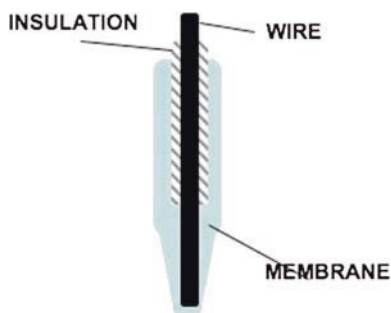
Fig. 6.18 Coated wire electrode (CWE)

6.2.3.2.1 Coated Wire Electrodes

A schematic diagram of a coated wire electrode is shown in Fig. 6.18. The internal wire is dip-coated with ion-selective polymeric membrane. It can be also pressed or glued to the solid-state membrane, thus forming a compact and inexpensive ion sensor (Freiser, 1980). Noble metals or graphite were used in the early designs of the CWE, mainly for their electrochemical stability. Such electrodes often exhibited unacceptable drift; the high value of the internal charge-transfer resistance between the membrane and the wire was the reason. It is not surprising, because typical electrode reactions at noble metals are of the redox type; that is, the charge-transferring species is the electron.

In contrast, in most ion-selective membranes the charge conduction is done by ions. Thus, a mismatch between the charge-transfer carriers can exist at the noble metal/membrane interface. This is particularly true for polymer-based membranes, which are invariably ionic conductors. On the other hand, solid-state membranes that exhibit mixed ionic and electronic conductivity such as chalcogenide glasses, perovskites, and silver halides and conducting polymers (Lewenstam and Hulanicky, 1990) form good contact with noble metals.

There have been other approaches to the solution of the problem of the internal contact. The most direct one involves the interposition of a thin layer of aqueous gel that contains the fixed concentration of the salt of the primary ion. This approach, which has met with only limited success, can be seen as an attempt to miniaturize the conventional ISE structure. The main reason for its failure is the fact that electrically neutral species can permeate through the membrane and reach finite activities inside each individual phase of the sensor structure. Namely, water permeating through the membrane reaches its osmotic equilibrium, according to the concentration of the solutes present in the inside gel layer. This can lead not only to the significant change of internal activity of the primary salt inside the gel, but often to a catastrophic failure of the whole structure when the osmotic pressure inside the gel exceeds the limits of the mechanical strength of the membrane. This problem is difficult to avoid because the concentration of otherwise neutral solutes in the sample is not known a priori and cannot be controlled during the measurement.

The second problem relates to the "Severinghaus-type interference," which is caused by penetration of electrically neutral chemical species with acid/base or redox properties through the membrane (e.g. acetic acid or iodine). (See Section 6.3.2 for more on Severinghaus electrodes.) They can then dissociate inside the gel, change its pH and/or the redox potential, and thus affect the interfacial potential (Fogt et al., 1995). In principle, this problem can be avoided by establishing a dominating electrochemical process with as high an exchange current density as possible at this interface. This process then controls the interfacial potential, which becomes independent of such interference. In principle, the same effect can cause problems in conventional ISE; however, due to the large volume (e.g. 3–5 mL) of the internal solution, such changes would take a long time to be noticeable and it does not happen in practice.

It has been found that a much more stable CWE can be made by replacing the noble metal with less noble conductors, such as copper or silver. In particular, the use of a thin coating of AgCl on top of the silver wire is possible. Coating the metal contact with an organic semiconductor also stabilizes the potential of the electrode. The reason is that organic semiconductors are mixed electronic/ionic conductors and thus can couple the electronic charge in the metal to the ionic charge in the membrane. An internal contact made of carbon should suffer from the same problems as the noble metal CWE. Such a contact exists in the so-called Selectrodes (Ruzicka and Lamm, 1971), in which the ion-selective membrane is mixed in with carbon paste. Carbon has similar electrochemical properties to those of noble metals. However, in Selectrodes the contact with the ISE membrane has a very large area that far exceeds the geometrical surface area of the electrode. Because the interfacial resistance is inversely proportional to the contact area, this interface is relatively stable.

The fluoride ion-selective electrode used in the cell discussed in Section 6.2.2.2.1 is one example of a nonsymmetrical potentiometric sensor. It cannot be properly called a "coated wire electrode" because the contact between the membrane and the internal connector has been made not by coating, but by some thin-film deposition technique (e.g. sputtering). Nevertheless, the electrochemistry involved in the operation of this sensor is quite the same. It is again essential to identify the charge-transferring species at each interface. In this case, solid AgF is used instead of the usual AgCl because fluoride ion is needed to couple the AgF and LaF_3 layers. The transformation from the ionic transport to the electronic one takes place at the Ag/AgF interface. The solid contact between the regular pH glass electrode and silver wire has been also made by using AgF as the intermediate ionic conductor (Fjeldly and Nagy, 1985) because fluoride ion has finite mobility in hydrated glass.

In summary, it is perfectly legitimate to design CWEs with solid-state internal contact and to expect electrochemical performance comparable to the conventional ISE. However, the design of the membrane/solid interface has to be done with the understanding of the electrochemical processes at such an interface. The problems include the drift scale with the surface area and the length of the internal contact, specifically with its parasitic capacitance and resistance. From this consideration alone, it can be concluded that such problems can be minimized by decreasing the
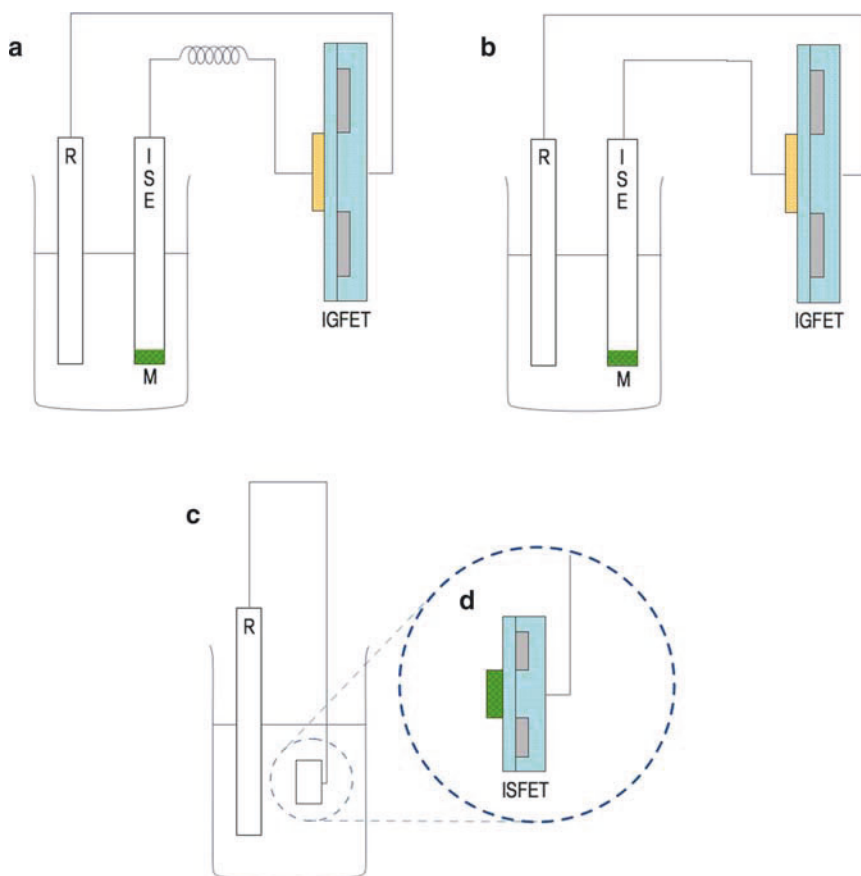
Fig. 6.19 Evolution of ion-sensitive field-effect transistor (ISFET) from insulated gate FET and ion-selective electrode (ISE)

distance between the ion-selective membrane and the amplifier input. Therefore, the next logical step in the evolution of solid-state sensors is either the hybrid ion sensor or the ion-sensitive field-effect transistor (Fig. 6.19).

The contact problems are mitigated in the hybrid ion sensor by making the internal conductor shorter and shorter until it is more natural to talk about its thickness rather than its length. The material of this internal contact has not changed during this transition and neither has the electrochemistry at the interface. Thus, the only difference between the coated wire and the hybrid sensor is the length (or thickness) of the contact. We therefore skip it and go directly to the solid-state ISFET, in which the thickness of the internal contact is zero. In other words, the ion-selective membrane is placed directly at the input dielectric of the field-effect transistor (Fig. 6.20).
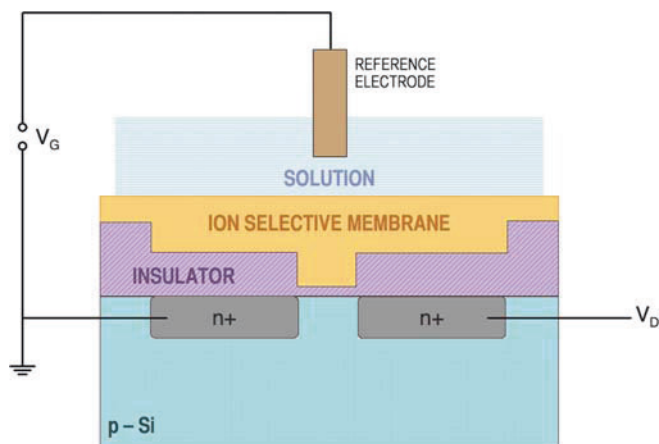
Fig. 6.20 Schematic of ion-sensitive field-effect transistor (ISFET)

6.2.3.2.2 Field-Effect Transistors

Field-Effect Transistors (FETs) are part of all modern pH meters. With the introduction of ion-sensitive field-effect transistors, they have both been brought to the attention of chemists. In order to understand the principles of operation of these new electrochemical devices, it is necessary to include the FET in the overall discussion of the electrochemical cell. The outline of the operation of an insulated gate field-effect transistor is given in Appendix C.

The idea of the ion-sensitive field-effect transistor was introduced independently by Bergveld (1970) and by Matsuo et al. (1971) in the early 1970s. The metal gate of an ordinary IGFET was removed and the gate insulator was directly exposed to the electrolyte solution. Such device proved to be sensitive to changes of pH. A few years later, an ion-selective membrane was placed directly over the gate dielectric (Moss et al., 1975) and an entire family of ISFETs sensitive to different ions was born (Janata and Huber, 1980). A schematic diagram of the ISFET is shown in Fig. 6.20.

The heart of an ISFET is the gate. The solid-state part of this sensor is the same as in an IGFET but the metal gate has been replaced with a chemically sensitive layer. The electrical path has been closed with a reference electrode and both have been placed in the conducting solution. The electronic parts of the device are protected by a suitable encapsulant. In addition to ion-selective membranes, other types of selective layers have been integrated with the IGFET core device, giving rise to a group of potentiometric sensors known as chemically sensitive field-effect transistors (CHEMFET; Janata, 2004). It is shown in Appendix C how the gate voltage V_G controls the drain current I_D in the transistor. It was thought initially (Bergveld, 1970) that ISFETs could be operated without a reference electrode. The idea attracted a lot of attention, because a reference electrode is a major inconvenience. However, such operation would violate the one-capacitor rule discussed earlier. In order to
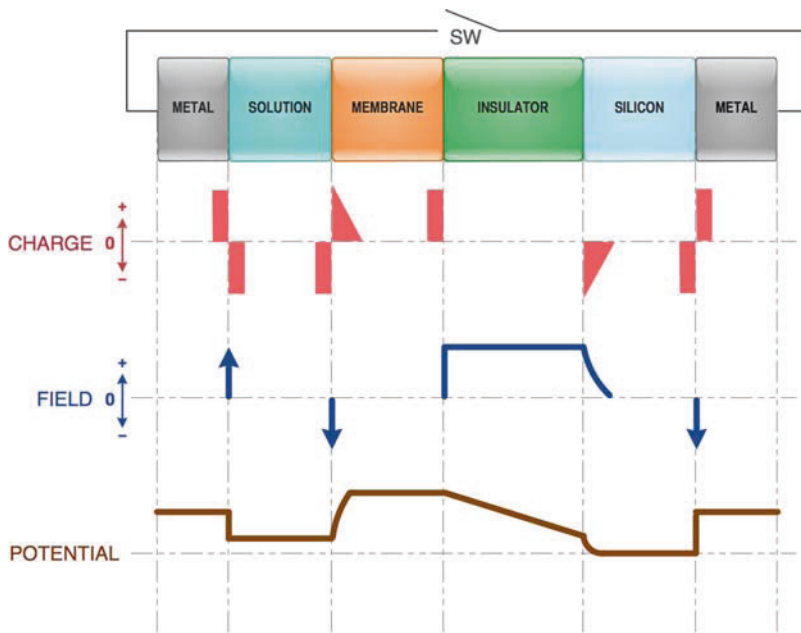
Fig. 6.21 Profiles of charge, electric field, and potential distribution through the gate of an ion-sensitive field-effect transistor (ISFET)

understand this argument, it is necessary to perform a charge-transfer analysis of the ISFET gate structure (Fig. 6.21).

This schematic represents the cross-section through the entire electrochemical cell, including the reference electrode. Out of this analysis will come the operating characteristics of potassium ion-selective field-effect transistor (K^+-ISFET). We use the reference electrode (1) of the first kind, without liquid junction, which is the simplest case. Let us assume that the solution (2) also contains a small amount of ions that can partition into the potassium-selective membrane (3). Therefore, the interface between the solution and the membrane is nonpolarized. A concrete example is a solution containing $AgNO_3$ and KNO_3. The insulator (4) is assumed to be ideal; that is, no charge can cross it and it is thicker than electron tunneling distance ($d > 10$ nm). Layer (5) is the semiconducting substrate of the transistor. For the sake of simplicity, the metal (6) is identical to metal (1) of the reference electrode. The switch S_w allows us to perform a Thought Experiment. It represents operation with (S_w closed), and without (S_w open) the reference electrode. Figure 6.21 indicates the charge, field, and potential profiles across this structure.

Note that this is a very simplified case. A liquid junction, dual-layer insulator, trapped charges in the insulator, surface states at the insulator/semiconductor interface, channel doping profile, and multiple connecting metals have been omitted, for the sake of simplicity. They would be present in all real devices and situations, but would not affect the "thought analysis" in any significant way.

From the thermodynamic point of view, this is a multiphase system for which, at equilibrium, the Gibbs equation (A.20) must apply at each interface. Because there is no charge transfer in and out of layer (4) (an ideal insulator) the sandwich of the layers (3)/(4)/(5) also represents an ideal capacitor. It follows from the Gibbs equation that this system will reach electrostatic equilibrium when the switch Sw is closed. On the other hand, if the switch Sw remains open, another capacitor (1)/(?)/(6) is formed, thus violating the one-capacitor rule. The "?" signifies the undefined nature of such a capacitor. The "open switch" situation is equivalent to operation without a reference electrode (or a signal return). Acceptable equilibrium electrostatic conditions would be reached only if the second capacitor had a defined and invariable geometry.

Let us now analyze Fig. 6.21. The starting point is the semiconductor (5) whose inner (Galvani) potential is

$$\varphi_5 = (\mu_5^e - \tilde{\mu}_5^e)/F \tag{6.54}$$

Here μ_5^e is the chemical potential of electrons in the semiconductor (the electron–lattice interaction energy) and $\tilde{\mu}_5^e$ is the electrochemical potential of an electron in phase 5, normally known as the Fermi level (see Appendix A for explanation of the difference). Similarly, the inner potential of the membrane (3) is

$$\varphi_3 = (\tilde{\mu}_3^i - \mu_3^i)/z_i F \tag{6.55}$$

where $\tilde{\mu}_3^i$ and μ_3^i are the electrochemical and chemical potentials of ion i in phase 3. The potential difference across the insulator and the semiconductor space charge region is then

$$\varphi_5 - \varphi_3 = (\mu_5^e - \tilde{\mu}_5^e)/F - (\tilde{\mu}_3^i - \mu_3^i)/z_i F \tag{6.56}$$

It is now necessary to specify the relationship between species i in the membrane (3) and the electron in the semiconductor (5). We know that ion i can transfer from the solution (2) into the membrane (3). Therefore, its electrochemical potential in the two phases must be the same.

$$\tilde{\mu}_3^i = \tilde{\mu}_2^i = \mu_2^i + z_i F \varphi_2 \tag{6.57}$$

where φ_2 is the inner potential of the solution. Similarly, the Fermi levels in the semiconductor and in the metal (6) are equalized and, because the metal (6) is the same as the metal (l, the reference electrode), we can write

$$\tilde{\mu}_5^e = \tilde{\mu}_1^e = \mu_1^e - F \varphi_1 \tag{6.58}$$

There is an equilibrium between the metal cations and electrons in the metal, expressed by $M_1 = M_1^+ + e_1^-$, and for which we can write

$$\mu_1^M = \mu_1^{M+} + \mu_1^e \tag{6.59}$$

On substituting for μ_1^e in (6.58) we obtain

$$\tilde{\mu}_5^e = \tilde{\mu}_1^M - \mu_1^{M+} - F \varphi_1 \tag{6.60}$$

If (6.56), (6.58), and (6.60) are combined and rearranged, then the potential difference across the insulator becomes

$$\varphi_5 - \varphi_3 = \frac{\mu_5^e - \mu_1^e}{F} - \frac{\mu_2^i - \mu_3^i}{z_i F} + \varphi_1 - \varphi_2 \tag{6.61}$$

The first term on the right-hand side of (6.61) is the contact potential between the semiconductor and the metal of the reference electrode (1):

$$\varphi_5 - \varphi_1 = \frac{\mu_5^e - \mu_1^e}{F} = \pi_{\text{cont.}} \tag{6.62}$$

The second term is the potential at the interface of the solution and the membrane, and is related to the activity of the ion in solution (2) by the Nernst equation.

$$\frac{\mu_2^i - \mu_3^i}{z_i F} = \pi_0' + \frac{\Re T}{z_i F} \ln a_2^i \tag{6.63}$$

The constant activity a_3^i in the membrane (3) is included in the term π_0'. The third term is the potential of the reference electrode $\varphi_1 - \varphi_2 = \pi_{\text{ref}}$. Equation (6.56) can now be written as

$$(\varphi_3 - \varphi_5) = \pi_{\text{cont.}} + \pi_0' + \pi_{\text{ref}} + \frac{\Re T}{z_i F} \ln a_2^i \tag{6.64}$$

It is the contribution of the electrochemical part of the IGFET to the overall gate voltage. Therefore, it can be added to the externally applied gate voltage V_G (C.25) yielding the complete equation for the ISFET. There are again two operating regimes, nonsaturation and saturation. In the linear region ($V_D < V_G - V_T$), the ISFET current is

$$I_D = \frac{\mu_n C_0 W}{L} V_D \left(V_G - V_T^* - \frac{V_D}{2} + \pi_{\text{ref}} + \frac{\Re T}{z_i F} \ln a_2^i \right) \tag{6.65}$$

In the saturation region ($V_D > V_G - V_T$) and

$$I_D = \frac{\mu_n C_0 W}{2L} \left(V_G - V_T^* + \pi_{\text{ref}} + \frac{\Re T}{z_i F} \ln a_2^i \right)^2 \tag{6.66}$$

The threshold voltage V_T^* of ISFET in both modes is defined as

$$V_T^* = V_T + \pi_{\text{cont.}} + \pi_0' \tag{6.67}$$

It contains parameters related to the solid-state (i.e., V_T) as well as to the chemical part (i.e., $\pi_{\text{cont.}}, \pi_0'$) of the sensor. Because the reference electrode is physically separate from the ISFET, its potential is not included in the threshold voltage. However, that choice is rather arbitrary. Whether the actual output follows (6.65) or (6.66) depends on the externally applied gate voltage V_G, which has nothing to do with the

electrochemistry at the interface. The measurement of current, utilizing (6.65) and (6.66), is done with the circuit shown in Fig. 4.28a.

Equations (6.65) and (6.66) relate the ISFET output to the activity in a rather awkward way and can be rearranged for "feedback" (constant current) operation. Thus, for an ISFET in the nonsaturation (linear) region, we have from (6.65)

$$V_G = \frac{LI_D}{W\mu_n C_0 V_D} + V_T^* + \frac{V_D}{2} - \pi_{ref} - \frac{\Re T}{z_i F} \ln a_i \qquad (6.68)$$

and for the saturation region, from (6.66)

$$V_G = \left(\frac{2LI_D}{W\mu_n C_0}\right)^{1/2} + V_T^* + \pi_{ref} + \frac{\Re T}{z_i F} \ln a_i \qquad (6.69a)$$

or when the Nikolskij–Eisenman equation (6.67) is used

$$V_G = \left(\frac{2LI_D}{W\mu_n C_0}\right)^{1/2} + V_T^* + \pi_{ref} + \frac{\Re T}{z_i F} \ln \left(a_i + \sum_j K_{i,j} a_j^{z_i/z_j}\right) \qquad (6.69b)$$

Equation (6.69b) implies that the sensor is operating in the background electrolyte containing interfering ions j. The output gate voltage in the absence of the primary ion is then given by (6.70), which contains all constant terms belonging to the transistor, the reference electrode potential, and the logarithmic term contributed by the interfering ions.

$$V_G^* = \left(\frac{2LI_D}{W\mu_n C_0}\right)^{1/2} + V_T^* + \pi_{ref} + \frac{\Re T}{z_i F} \ln \sum_j K_{i,j} a_j^{z_i/z_j} \qquad (6.70)$$

The operating point of the ISFET is set by the value of the drain current ID. Subtraction of (6.70) from (6.69b) then gives the response of the ISFET to the primary ion in the given electrolyte, which is probably the most useful analytical information. It is usually reported as the "ISFET response" ΔV_G.

$$\Delta V_G = \frac{2.303 RT}{z_i F} \log \left(\frac{a_i}{\sum_j K_{i,j} a_j^{z_i/z_j}} + 1\right) \qquad (6.71)$$

As can be seen from Fig. D.8 (Appendix D), ISFET is more sensitive to changes of V_G when operated in the saturation region, hence also more sensitive to changes of the interfacial potential at the membrane/solution interface. Therefore, the feedback saturation mode is preferable.

Because all except the last terms on the right-hand side of (6.68) and (6.69) are constant, we have a simple functional relationship between the output of the ISFET and the activity of the ion in solution. From a practical point of view, an operational amplifier in a source–follower configuration, as shown in Fig. 6.22b can be used.
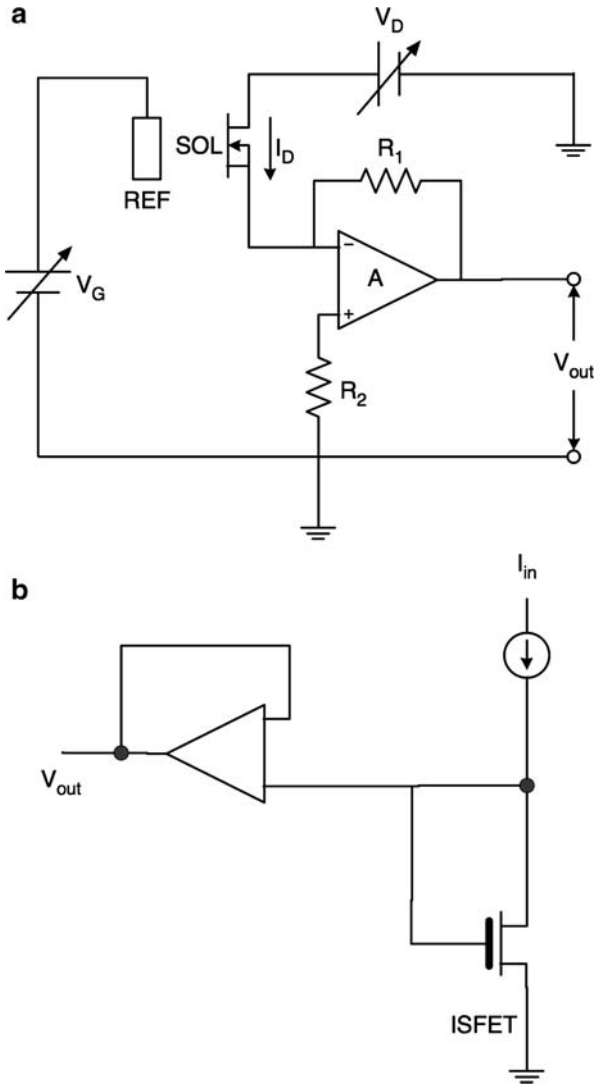
Fig. 6.22 Analog circuits for operation of ion-sensitive field-effect transistor (ISFET) (**a**) in constant applied voltage mode ((6.62) and (6.63)) and (**b**) in (source–follower) constant current feedback mode ((6.65) and (6.69))

The operating point of the ISFET is set and held constant by the drain current I_D. Any change in the potential difference between the solution and the ion-selective membrane due to a change in ionic a_i appears as a change in the V_G((6.69) through (6.71)). Naturally, it is possible to realize the feedback circuit in many different ways, either in analog or in digital form. Because nothing else changes, the ISFET readout is in mV, as is customary in conventional potentiometric sensors.

pH ISFET

The close integration of the selective layer with the amplifier, as it is done in ISFET, offers some unique sensing possibilities. We must remember that the field-effect transistor is basically a charge-sensing device. From the very beginning (Bergveld, 1970; Matsuo et al., 1971) of development of this device it was noted that ISFET with a bare gate insulator responds to changes of pH when exposed to the solution. Because pH is such an important parameter, considerable effort has been devoted to the explanation of the mechanism of this response. The slope of this response has been found to vary between 50% (for SiO_2) to 92% (for silicon nitride) of the theoretical value. Other materials, oxides such as Al_2O_3 or Ta_2O_5, have also been reported to yield a response in this range. Moreover, the response appeared to depend also on the method of preparation. Clearly, there is a pH-dependent charge at the solution/transistor interface. It is necessary to pause here and to realize that all these materials are very good bulk insulators. Therefore, they cannot be used as pH-sensing membranes in a conventional ISE configuration because they would create another capacitor in the circuit. However, in an ISFET configuration, these materials are part of the input capacitor and their use as a sensitive layer is justified.

It would be possible to accept the fact that the pH-dependent charge resides "somewhere at the insulator surface" and that the corresponding image charge in the semiconductor affects the output. However, such a superficial assessment could lead and has led to some erroneous conclusions, namely in the development of charge-related biosensors. Let us remember the basic difference between the nonpolarized (resistive) and polarized (capacitive) interface (Section 5.3). The argument about the location of the pH-dependent charge has revolved around this distinction. One school of thought has been that the charge is located in one plane and that the interface behaves as a capacitor at which the charge is generated by the deprotonation/protonation of the surface-bound hydroxyl sites in that plane. This model has its origin in some work on oxide colloids and it is the basis of "Site-Binding Theory" (SBT; Schenck, 1977). In this theory, it is assumed that there are ionizable binding sites present at the surface of the insulator that determine the distribution of the compensating charge in the adjacent layer of the electrolyte. This distribution is the result of the combination of the interplay between the thermal and electrostatic forces originating from the pH-dependent charge at the surface. The exponential decay of the potential is again governed by the Poisson–Boltzmann distribution, as was the case in the ideally polarized interface (Section 5.3). The decay of potential φ_x n the x-direction from the surface is exponential

$$\varphi_x = \varphi_0 \exp(-\kappa_D x) \tag{6.72}$$

where φ_0 is the surface potential due to the density of the surface charge. The effective thickness κ_D^{-1} of this space charge is given by (6.53); it depends on the dielectric constant ε, and on the bulk concentration C of the electrolyte. Possible complications arising from the presence of image charges at the electrolyte/dielectric

interface (Williams, 1975) are usually neglected. This model has been adopted and used mostly by electrical engineers.

The other model, preferred by electroanalytical chemists, postulates the existence of a hydrated layer of finite thickness, which contains the deprotonated/protonated sites. The creation of the boundary potential follows the argument similar to the one presented for the glass electrode. Clearly, in the limit of zero thickness of the hydrated layer, the two models merge. In the SBT model the interface acts as a capacitor with one plate located at the surface and the other at the average distance κ_D^{-1} in the solution. In other words, the interface is considered to be pH-sensitive, yet ideally polarized. On the other hand, the hydration layer model allows penetration of at least one type of ion (namely of hydrated proton) through the interface, and thus presents this interface as nonpolarized; how "nonpolarized" depends on the value of the exchange current density of the proton. The equivalent electrical model is again a parallel capacitor/resistor combination. From the point of view of the response, it is obvious that the interface described by the capacitive model (SBT) would respond to adsorption of any charged species inside the space charge region, as well as to the motion of the electrolyte. On the other hand, adsorption would have little or no effect on the potential difference at the interface described by the hydration layer model because that potential difference is uniquely and unequivocally given by the transfer of the dominating ion.

The common ground that accommodates both models has been proposed by Sandifer (1988) who has shown that the sub-Nernstian response of some materials and the resulting adsorption and stirring effects can be rationalized by considering the number of available binding sites and the thickness of the hydrated layer. The magnitude of the exchange current density depends on the concentration of the binding sites inside the hydrated layer. For the glass electrode, this is estimated to be 3.2 M (Eisenman, 1969). As the number of binding sites decreases from 3 M to 0.1 M, the interfacial potential of a 100 nm thick layer is increasingly affected by adsorption (Fig. 6.23).

At the same time, the response deviates more from the theoretical value the fewer proton-binding sites are present. On the other hand, both the adsorption effects and the sub-Nernstian behavior vanish if the thickness of the hydrated layer is allowed to increase up to 800 nm (Fig. 6.24). It is seen from this model that as the thickness of the hydrated layer exceeds the thickness of the space the adsorption effects and the sub-Nernstian behavior disappear.

This model goes a long way towards explaining most experimental results reported in the literature for ISFETs with oxide or nitride surfaces. Unfortunately, the properties of these materials prepared in different laboratories are very different. It is known that silicon nitride (really silicon oxynitride $Si_3N_4O_x$) forms an oxygen-rich layer at the surface, whose thickness depends on the deposition conditions. This "passivation" layer forms rapidly (in a matter of hours) and is very stable, even under continuous exposure to aqueous electrolyte. The hydration of this layer seems to fit the requirements and predicted behavior of the Sandifer model. Consequently, ISFETs that have been exposed to aqueous solution for more than one hour show no adsorption effects which would be expected from the SBT model.
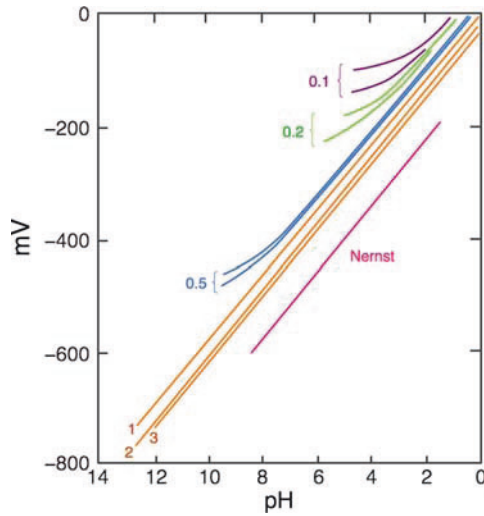
Fig. 6.23 Dependence of the Nernst potential on the density of the proton-binding sites and on adsorption. Upper curves within the bracketed sets correspond to the absence of adsorption. The lower curves show the effect of 100 mM charged adsorbate (adapted from Sandifer 1998)
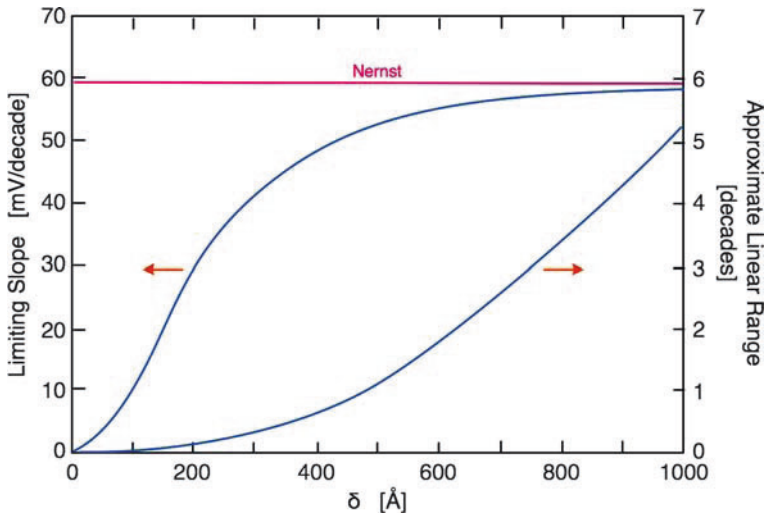
Fig. 6.24 Dependence of Nernst response of the thickness of the hydrated layer (adapted from Sandifer, 1998)

The hydration layer model more or less fits the behavior of all oxides and of silicon oxynitride. The latter has a great advantage over other pH-sensitive layers because it is an excellent passivation material itself that can be prepared virtually pinhole-free. On the other hand, even thermally grown SiO_2 has pinholes and cracks that form a leakage path. It is that leakage path that apparently served as a

"pseudo-reference electrode" and led to the erroneous conclusion that ISFETs can be operated without a reference electrode. It is rather remarkable that nitride was used in the first successful pH ISFET (Matsuo et al., 1971), whereas the credit for the introduction of ISFET usually goes to the nonfunctioning silicon dioxide version responsible for experimental artifacts (Bergveld, 1970).

ISFETs for Other Ions

The development of ISFETs for other ions followed the path of macroscopic ISEs. Various ISE membranes have been synthesized and solvent cast on the gate dielectric (Janata, 2004). By the end of the 1980s, ISFETs for every ion measurable by the ISEs had been developed and published. It was emphasized in the introduction to asymmetric membrane ion sensors (Section 6.2.3.2) that the proper design of the solid/membrane interface is critical for proper functioning of these devices. As always, the dominating consideration has been the magnitude of the exchange current density. If it is low this interface becomes capacitive or "quasi-capacitive" and the one-capacitor rule is violated. It usually results in instability of the device. In the case of ISFET, the ion-selective membrane is deposited directly over the one and only input capacitor. Therefore, no charge transfer happens at such an interface and consideration of exchange current density is a moot point. In other words, the electric field that is responsible for the semiconductor field effect and that modulates the conductivity of the ISFET originates at the membrane/insulator interface.

In ISFETS utilizing polymeric ion-selective membranes, it has been always assumed that these membranes are hydrophobic. Although they reject ions other than those for which they are designed to be selective, polymeric membranes allow permeation of electrically neutral species. Thus, it has been found that water penetrates into and through these membranes and forms a nonuniform concentration gradient just inside the polymer/solution interface (Li et al., 1996). This finding has set the practical limits on the minimum optimal thickness of ion-selective membranes on ISFETS. For most ISE membranes, that thickness is between 50–100 μm. It also raises the issue of optimization of selectivity coefficients, because a partially hydrated selective layer is expected to have very different interactions with ions of different solvation energies.

Permeation through the membrane of electrically neutral molecules also underscores two other important practical problems. Ingress of water may compromise the adhesion of the polymeric membrane with the gate dielectric. It has also been observed that some otherwise neutral gases, such as CO_2, NH_3, and SO_2, may hydrolyze and change the local pH at the interface of the hydrated layer of the dielectric and the membrane. If the natural buffer capacity of that interface is exceeded the ISFET may respond to the sustained, high partial pressure of such gases (Fogt et al., 1985). The ISFET, in essence, begins to work as a miniature Severinghaus electrode. (We discuss these in Section 6.3.2.) It needs to be pointed out that such effects can be achieved at any conventional macroscopic electrode with internal filling solution. It is only a matter of time of exposure and concentration of the gas.

In other words, it proves that it is possible to design any experiment in such a way that it exceeds the limit of its validity. Both the adhesion and the Severinghaus effect can be mitigated by proper design of the membrane (Li et al., 1988) or by chemical modification of the surface of the dielectric.

Multiple Ion ISFETs

Solid-state chemical sensors based on silicon technology can easily accommodate multiple sensing channels as well as intergated signal-processing circuitry. This advantage ultimately leads to the fabrication of large chemical sensing arrays that far exceed the information acquisition capability of individual sensors (Chapter 10). An example of this trend is the dual-channel ISFET, which allows simultaneous sensing of two ions. Figure 6.25 shows a schematic of ISFET in which one channel is made sensitive to sodium ion activity and the other to the chloride ion activity (Bezegh et al., 1987).

When two identical transistors, one with the cation-sensitive membrane (+) and the other with the anion-sensitive membrane (–), are operated in the feedback mode the individual responses are given by (6.68), which for the Na^+ ISFET is

$$V_G^{Na^+} = \left(\frac{2LI_D}{W\mu_n C_0} \right)^{1/2} + V_T^* - \pi_{ref} + \frac{2.3\Re T}{F} \log a_{Na^+} \qquad (6.73)$$

and for the chloride ion

$$V_G^{Cl^-} = \left(\frac{2LI_D}{W\mu_n C_0} \right)^{1/2} + V_T^* - \pi_{ref} - \frac{2.3\Re T}{F} \log a_{Cl^-} \qquad (6.74)$$
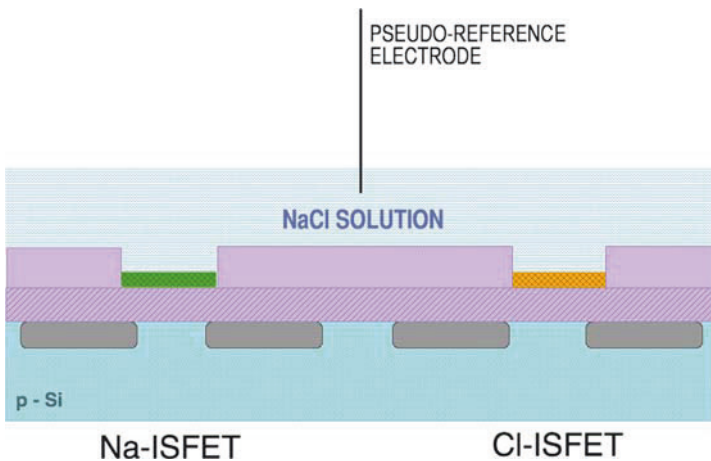
Fig. 6.25 Dual ion-sensitive field-effect transistor (ISFET) for measurement of sodium chloride concentration
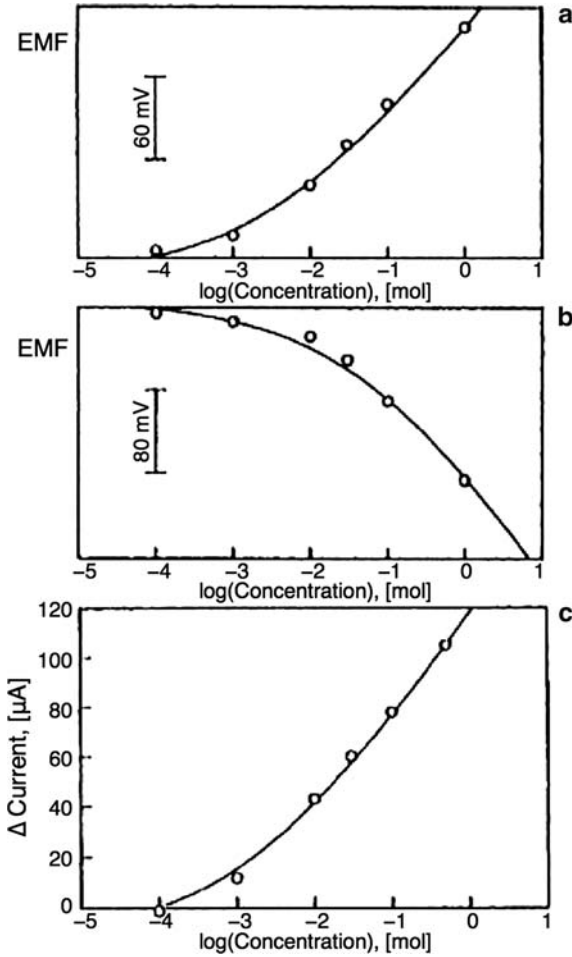
Fig. 6.26 Individual and differential response of the NaCl ion-sensitive field-effect transistor (ISFET) sensor (**a**) response of Na-ISFET and (**b**) Cl ISFET both measured against regular reference electrode; (**c**) differential current measurement of concentration of NaCl in 0.01 M MgSO$_4$ solution (adapted from Bezegh et al., 1987)

Because the two ISFETs are on the same chip, and the reference electrode is common to both, subtraction of (6.72) from (6.71) yields the value of activity of sodium chloride *salt* with double the slope (Fig. 6.26).

$$\Delta V_G = V_G^{Na^+} - V_G^{Cl^-} = \text{const.} + 2x \frac{2.3\Re T}{F} \log a_{NaCl} \qquad (6.75)$$

This probe has been fabricated and used for the determination of sodium chloride in perspiration, as with the test for detection of cystic fibrosis (Bezegh et al., 1988) The constant in (6.73) accounts for nonideality and differences in the two ISFETs.

Another remarkable feature of that probe is its complete independence of the reference electrode potential and nearly perfect temperature compensation. Thus, a pseudo-reference such as silver wire, functioning only as a signal return, can be used.

6.3 Potentiometric Sensors for Neutral Species

At first glance, this title seems to be a misnomer. After all, a potential is always generated by separation of charge, for example, by the partitioning of ions at the selective membrane/solution interface. Nevertheless, it is correct. Even potentiometric sensors for ions really respond to change of activity of a neutral substance, a salt, or rather to only one part of it, the cation or the anion. In the sensors discussed in this section, the electrically neutral substrates give rise to charged species by some secondary mechanism, such as their catalytic decomposition into ions, hydrolysis of acidobasic gas, or by formation of the charge-transfer complex with a neutral dopant in the semiconductor. They all end up creating a potential change that is then measured, at zero current, against some reference potential.

6.3.1 Potentiometric Enzyme Sensors

By the transduction mechanism, these sensors belong to the zero-flux boundary group, together with enzymatic thermal and enzymatic optical sensors. In fact, the verification of the proposed model which was discussed in Chap. 3 has been done on enzymatic field-effect transistors for the diffusion–reaction mechanism involving the oxidation of β-D-glucose catalyzed by the glucose oxidase (GOD)/catalase system (Caras et al., 1987) In both cases, hydrogen ion is the species detected by the transistor surface. The experimental and calculated responses according to the above model are shown in Figs. 1.12–1.14 (penicillin example) and Figs. 1.17–1.19 (GOD example). As has been pointed out earlier, their agreement verifies the validity of the model and allows us to construct the concentration profiles inside the gel layer. It is quite obvious that they are not linear, either in the steady-state or during the transition period. This shows that the early models of potentiometric enzymatic sensors were not correct. It is also not surprising to find that these devices have widely different lifetime, time response, detection limits, and sensitivity. The value of the experimentally verified model is then mainly in establishing the design parameters, not only for potentiometric but also for other zero-flux boundary enzymatic sensors.

Because enzymes present such an attractive possibility for achieving chemical selectivity, enzyme electrodes were the first enzymatic chemical sensors (or first biosensors) made. The early designs used any available method of immobilization of the enzyme at the surface of the electrode. Thus, physical entrapment using dialysis membranes, meshes, and various covalent immobilization schemes have been
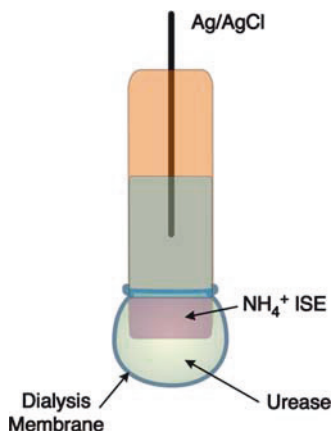
Fig. 6.27 Schematic of potentiometric enzyme electrode for urea, based on ammonium ion-selective electrode

employed (Section 1.2.4). The enzyme-containing layer is then simply added onto the existing electrochemical sensor (Fig. 6.27). The use of ISFETs enables precise control over the geometry of the enzyme layers as well as the miniaturization of these devices.

The choice of the ion sensor clearly depends on the type of the enzymatic reaction, namely on the products and reagents of that reaction and on the conditions of the sample. Thus, for example, there are several possibilities of the choice of the ion sensor for enzyme electrode for urea.

$$NH_2 - CO - NH_2 + H_2O \overset{urease}{\Rightarrow} 2NH_3 + CO_2$$
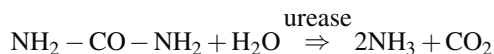
One could immobilize the urease layer on top of a Severinghaus electrode for CO_2 or NH_3 (Section 6.3.2) and use the device as an enzymatic–potentiometric gas sensor. The primary disadvantage of such an arrangement would be its slow response time. A more direct way is through the detection of the ionic species resulting from the hydrolysis of ammonia and carbon dioxide.

$$NH_3 + H_2O \Leftrightarrow NH_4^+ + OH^-$$

and

$$CO_2 + H_2O \Leftrightarrow HCO_3^- + H^+$$

An ionic sensor for the ions involved in the above reactions can be used. The choice is usually dictated by the conditions of the sample. Thus, at low pH, the species of choice would be NH_4^+ and at high pH it would be HCO_3^-. For the neutral pH range the best performance would be obtained from monitoring the changes of the pH itself.

Table 6.5 Example of potentiometric enzyme electrodes

Substrate	Enzyme	Sensor	Stability	Response	Range [M]
Urea	Urease	NH_4 + ISE	>4 mo	1–2 min	10^{-2}–10^{-4}
		pH ISE	3 w	5–10 min	10^{-3}–10^{-5}
		Gas (NH_3)	>4 mo	2–4 min	10^{-2}–10^{-5}
		Gas (CO_2)	3 w	1–2 min	10^{-2}–10^{-4}
Glucose	Glucose Oxidase	pH ISE	1 w	5–10 min	10^{-1}–10^{-3}
L-Amino Acids	L-AA Oxidase	NH_4 + ISE	>1 mo	1–3 min	10^{-2}–10^{-4}
L-Tyrosine	L-Tyrosine Decarboxylase	Gas (CO_2)	3 w	1–2 min	10^{-1}–10^{-4}
L-Glutamine	Glutaminase	NH_4 + ISE	2 d	1 min	10^{-1}–10^{-4}
L-Asparagine	Asparaginase	NH_4 + ISE	1 mo	1 min	10^{-2}–10^{-4}
D-Amino Acids	D-AA Oxidase	NH_4 + ISE	1 mo	1–3 min	10^{-2}–10^{-4}
Penicillin	Penicillinase	pH ISE	1–2 w	30–120 s	10^{-2}–10^{-4}
Amygdalin	β-Glucosidase	CN-ISE	3 d	10–20 min	10^{-2}–10^{-5}
Nitrite	NO_2 reductase	Gas (NH_3)	3–4 mo	2–3 min	10^{-2}–10^{-4}

Because each enzyme sensor has its own unique response, it is necessary to construct the calibration curve for each sensor separately. In other words, there is no general "theoretical" response relationship, in the same sense as the Nernst equation is. As always, the best way to reduce interferences is to use two sensors and measure them differentially. Thus, it is possible to prepare two identical enzyme sensors and either omit or deactivate the enzyme in one of them. This sensor then acts as a "reference." If the calibration curve is constructed by plotting the difference of the two outputs as the function of concentration of the substrate, the effects of variations in the composition of the sample as well as temperature and light variations can be substantially reduced. Examples of potentiometric enzyme electrodes are listed in Table 6.5.

Despite the large amount of research done on these devices and the relatively detailed understanding of their operation, they have not yet found widespread use. There are several reasons. The main one seems to be the insufficient lifetime (both "shelf" and "in-use"), due to the degradation of the ex vivo enzyme preparations. Particularly when they are used in real samples, various reversible and irreversible inhibitors limit their usable lifetime. This problem seems to be mitigated when the enzymes are used in their "natural" environment, in cell cultures, tissue slices, and even in whole organs (Arnold, 1986).

The second limitation comes from the fact that, despite the exquisite selectivity of enzymatic reactions, the selectivity of the potentiometric enzyme sensors is very poor. The reason for this lies in the detection mechanism itself, when factors such as the buffer capacity of the sample seriously interfere.

6.3.2 Severinghaus Electrodes

The original design was developed specifically for sensing of carbon dioxide (Severinghaus, 1965), but the principle on which these gas sensors operate is general. The key element is a small compartment into which the gas can penetrate through a semipermeable membrane (Fig. 6.28).

The gas undergoes a chemical transformation, for example, hydrolysis, in which some detectable ion or ions are produced. Thus, these sensors are based on measurement of localized changes in ion activity. The process can be described by the series of equilibria, the first being the partitioning of the gas between the sample and the electrode. The basic principle that applies to other gases is explained on the operation of the carbon dioxide sensor, the original Severinghaus electrode. The solubility of CO_2 in water is given by the equilibrium:

$$a_{CO_2} = SP_{CO_2} \tag{6.76}$$

where S is the solubility coefficient and P_{CO_2} is the partial pressure of CO_2. Hydrolysis inside the electrode proceeds according to the equation

$$CO_2 + H_2O \overset{K_a}{\Leftrightarrow} HCO_3^- + H^+$$
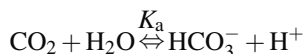
in which K_a is the acid dissociation constant.

$$K_a = \frac{a_{H^+} a_{HCO_3^-}}{a_{CO_2}} \tag{6.77}$$

In this case, an internal pH sensor is the choice. Because 1 mol of CO_2 yields 1 mol each of hydrogen and bicarbonate ion (6.77), after substitution from (6.76), becomes

$$K_a' = \frac{a_{H^+}^2}{SP_{CO_2}} \tag{6.78}$$
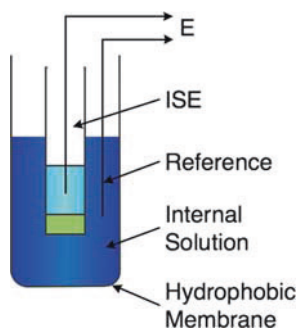
Fig. 6.28 Schematic of general Severinghaus electrode for sensing of hydrolyzable gases

Taking the negative logarithm and rearranging then yields

$$pH = \text{const.} - 0.5 \log P_{CO_2} \qquad (6.79)$$

where the $const. = 0.5 \, pK_a \log S$. The sensitivity of this electrode can be doubled if the internal filling solution contains a high concentration of bicarbonate (e.g., $0.1 \, M \, NaHCO_3$). In that case, the increase of the bicarbonate ion activity due to dissolved CO_2 is negligible compared to the total concentration and (6.77) becomes

$$K_a' = \frac{a_{H^+}}{\alpha P_{CO_2}} \qquad (6.80)$$

which then rearranges to

$$pH = \text{const.}' - \log P_{CO_2} \qquad (6.81)$$

It is important to note that if the detected species is hydrogen ion, then all acidobasic species are mutually interfering. Improved selectivity can be obtained by the judicial choice of the internal potentiometric element and to some extent by the selective permeability of the hydrophobic membrane. Thus, for example, for selective detection of HCN ($pK_a = 3.32$) the internal element should be a potentiometric sensor selective to CN^- and the pH of the internal electrolyte should be at least two pH units above the pK_a value (e.g., pH > 5.5). In that case, practically all HCN is dissociated.

Because the electrical circuit is closed inside the sensor, no "external reference electrode" is necessary and the Severinghaus-type electrode can be used for measurement in either gaseous or liquid samples. It is important to remember, however, that the potential of the internal reference electrode must remain constant. In principle, it would be possible to use a liquid junction but it would add to the complexity of the design. Because the counterion resulting from the dissociation equilibrium is the only interfering ion, and because it is present in a very low concentration, it is possible to ascertain the constancy of the reference potential by careful choice of the internal electrolyte. Thus, for example, in the CO_2 electrode the internal electrolyte is $0.1 \, M \, NaHCO_3$ and $0.1 \, M \, NaCl$ and Ag/AgCl is used as an internal reference element.

The performance characteristics of several Severinghaus-type electrodes are given in Table 6.6. The general facets are their simplicity of construction and of the operating principle. Because partitioning between the sample and the relatively large internal volume as well as dissolution equilibria are involved, the time constant of the Severinghaus electrode is typically 10–20 s. The addition of gas from the outside to the inside of the electrode can be viewed as a "titration" of the internal volume. For that reason, the dynamic range depends on the concentration of the internal electrolyte. Clearly, the lower the bicarbonate concentration is, the wider the dynamic range. However, for very low concentrations, the sensitivity of the electrode (slope) will decrease due to the transition from the response given by (6.81) to that given by (6.79).

Table 6.6 Performance characteristics of Severinghaus-type gas electrodes

Species sensed	Sensor	Internal electrolyte	Lower limit(M)	slope	Sample preparation	Interferences
CO_2	H^+	$0.01\,M\,NaHCO_3$	10^{-5}	$+60$	$<pH4$	
NH_3	H^+	$0.01\,M\,NH_4Cl$	10^{-6}	-60	$<pH11$	Volatile amines
Et_2NH	H^+	$0.1\,M\,Et_2NH_2Cl$	10^{-5}	-60	$<pH11$	NH_3
SO_2	H^+	$0.01\,M\,NaHSO_3$	10^{-6}	$+60$	HSO_4 buffer	Cl_2, NO_2 must be destroyed (N_2H_4)
	H^+	$0.1\,M\,NaHSO_3$	10^{-4}			
NO_2	H^+	$0.02\,M\,NaNO_2$	5×10^{-7}	$+60$	Citrate buffer	SO_2 must be destroyed (CrO_4^{2-}) CO_2 interferes
H_2S	S^{2-}	Citrate buffer (pH5)	10^{-8}	-30	$<pH5$	O_2 (ascorbic acid must be added to samples)
HCN	Ag^+	$KAg(CN)_2$	10^{-7}	-120	$<pH7$	H_2S (add Pb^{2+})
HF	F^-	$1\,M\,H^+$	10^{-3}	-60	$<pH2$	
$HOAc$	H^+	$0.1\,M\,NaOAc$	10^{-3}	$+60$	$<pH2$	
$2\,Cl_2$	Cl^-	H_2SO_4 buffer	5×10^{-3}	-60	$<pH2$	

Severinghaus electrodes have found wide application in clinical analysis. It is pertinent to mention here that the general principle of permeation of the gas through a hydrophobic membrane followed by its detection (with or without its solvolysis) has been used with different types of internal sensors, for example, optical, amperometric, conductimetric, or a mass sensor. The choice of the internal sensing element depends on the circumstances of the application in which the gas sensor would be used, such as the required time response, selectivity considerations, complexity of instrumentation, and so on.

6.3.3 Work Function Sensors

Work Function (WF) plays a key role in the physics and chemistry of materials. Phenomena such as the semiconductor field effect, photo- and thermionic electron emission (Allen and Gobelli, 1962), catalysis (Vayenas et al., 1996), and the like are dominated by the WF. This fundamental property of electronic materials is defined as "the minimum work required to extract an electron from the Fermi level E_F of a conducting phase, through the surface and place it in vacuum just outside the reach of the electrostatic forces of that phase" (Trasatti and Parsons, 1986). The reference level for this transfer is thus called the vacuum reference level. Because even a clean surface is a physical discontinuity, a surface dipole η with its associated electric field always appears at the surface of the condensed phase. Thus, the work of extracting the electron can be conceptually divided between the work required to

free the electron from the bulk and the work associated with its passage through the surface.

The energy with which electrons are bound in conducting materials is known as the electron affinity of the material. Materials with a high electron affinity bind electrons strongly and exhibit "noble" behavior (i.e., are relatively inert and do not oxidize spontaneously in air). Gold is an example. On the other hand, metals such as aluminum or copper are less noble and their surfaces, once exposed to air, are readily oxidized. When two dissimilar electronic conductors are placed in contact with each other, electrons flow from the material that is less noble (e.g., copper) to the more noble material (e.g., palladium) until an equilibrium is reached and the contact potential is formed at their junction. Because of the multitude of possible combinations of conductors in the real world, contact potential is the most ubiquitous of all junction potentials.

6.3.3.1 Kelvin Probe

Let us make a simple electrical circuit from palladium and copper. First the ends of the two metals are arranged in such a way that they form a parallel plate capacitor (Fig. 6.29). An electric field appears between the two plates, which is the result of the spontaneous separation of electrons, driven by the difference of their electron affinities.

In the case of a copper/palladium junction, electrons are withdrawn from the copper plate which becomes positively charged. They are then deposited in the palladium plate which becomes negative. A galvanometer and a battery are also placed in series to complete the circuit. It can be shown that the field inside the gap depends only on the materials forming the capacitor and not on material(s) connecting the two plates.

The two plates are vibrated; that is, the distance between them is periodically changed. In that case, the galvanometer registers AC current of the frequency of the mechanical vibration. Next, a compensating voltage from the variable voltage source is applied until the electric field is nulled out and the galvanometer registers zero current. At this point, the compensating voltage equals the WF difference
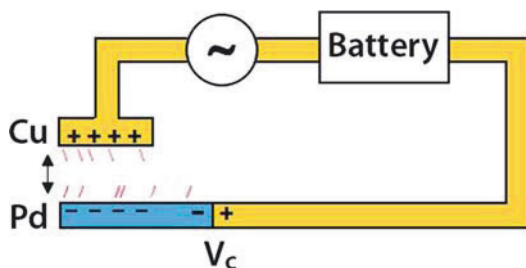
Fig. 6.29 Schematic of the Kelvin probe (vibrating capacitor)

divided by the electronic charge. If we now introduce hydrogen, we observe that a new value of compensating voltage is needed to null out the AC current. The reason is that hydrogen dissolves in Pd, forming palladium hydride, which has a lower work function. As a result, some electrons are returned to copper.

This simple and useful instrument is named the "Kelvin probe," after Lord Kelvin who introduced the concept of the contact potential in 1897 (Lord Kelvin, 1898) and demonstrated it by using a leaf electrometer. However, the first experiments of this kind were apparently made by Pellat, 16 years earlier (Pellat, 1881). The modern version of the vibrating plate capacitor was introduced by Zisman in the mid-1930s. Since then, it has been used widely by physicists, material scientists, and (to some extent) by physical chemists. Its measurements are referred to by a confusing variety of names, such as measurement of "contact potential", "surface potential", or "work function" of materials. It was first used for genuinely analytical purposes by G. Philips (1951) as a rather nonspecific detector for gas chromatography. Arrival of flame-ionization and other detectors eclipsed that use of the KP as an analytical instrument. The Pd-based IGFET was the first solid-state WF sensor (Lundstrom et al., 1975).

6.3.3.2 Suspended Gate Field-Effect Transistor

The Kelvin probe inspired the design of another solid-state device called the Suspended Gate Field-Effect Transistor (SGFET) (Blackburn et al., 1983). It resembles the Kelvin probe in that the gate conductor is suspended approximately $1\,\mu$m above the gate insulator, thus forming a gap of a capacitor (Fig. 6.30).
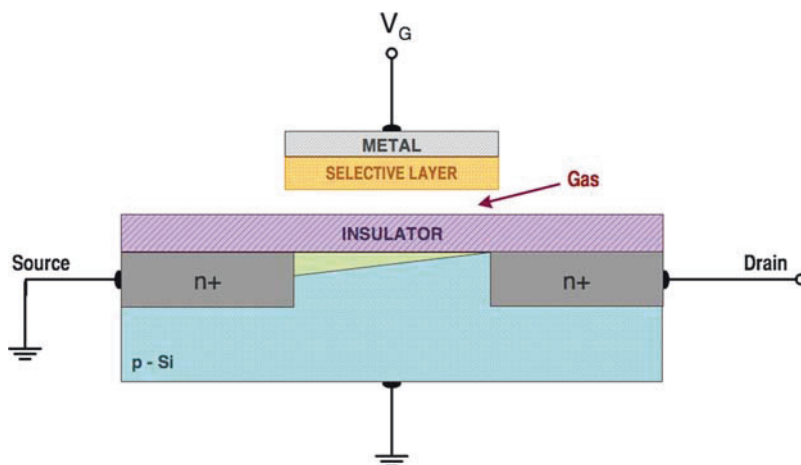
Fig. 6.30 Schematic diagram of the suspended gate field-effect transistor (SGFET). The selective layer is typically deposited electrochemically

However, unlike the KP, this bridge is not vibrated. Instead, the charge on the capacitor is determined from the drain current flowing through the semiconducting channel. With some difficulty it is possible to electrochemically deposit a selective layer under this suspended metal bridge and then modulate this layer chemically. Suspended gate field-effect transistors selectively sensitive to alcohol vapors, hydrogen cyanide, tributyl phosphate, hydrogen, and ammonia have been described. The one (and possibly the only) advantage of the SGFET is that the bridge can be operated at variable temperature (Zhang, 1995). However, its fabrication is quite complex and may not outweigh the advantage of variable temperature (Josowicz and Janata, 1988).

6.3.3.3 Field-Effect Transistors with Semiconductor Gate

A more direct approach is to deposit the electronically conducting, gas-permeable, and selective material directly over the gate insulator. The field-effect transistor with Pd gate, introduced by Lundstrom et al. (1975), is a perfect example. It is the most fundamental CHEMFET because it directly combines the chemical sensitivity and selectivity with IGFET operation. It can be electrically modulated as a normal transistor and all the equations that describe its operation (Appendix C) apply. The response to hydrogen is due to the modulation of the electron work function of the palladium gate and therefore of the threshold voltage. Although it establishes the basic principle, its limitation is that it is applicable only to hydrogen and to hydrogen-producing species.

The first report of a WF field-effect transistor with a gate material other than Pd apparently appeared in the late 1980s. A phthalocyanine film has been evaporated over the gate insulator and over the Al contacts (Burr et al., 1987). Since then, a variety of conducting polymers has been deposited over the transistor gate and evaluated as materials for gas sensing (Liess et al., 1996) (Fig. 6.31). The general
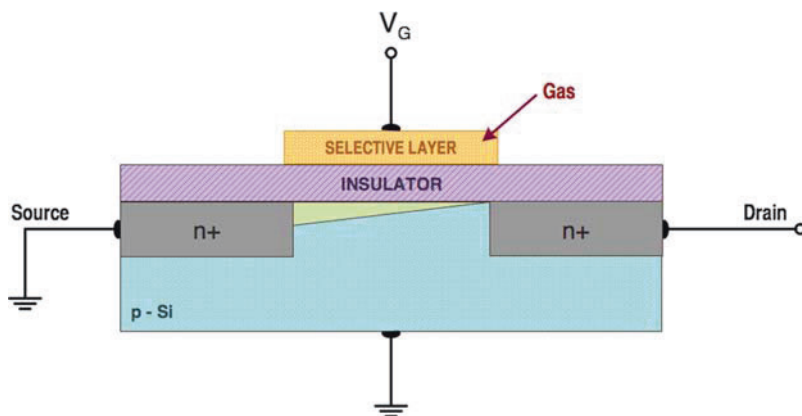
Fig. 6.31 Insulated gate field-effect transistor (IGFET) with electronically conducting, chemically selective gate

idea has been the same: modulation of the work function of the host matrix by the guest molecule. Both the intrinsic selectivity of the matrix and the selectivity created by the added specific binding sites (Janata and Josowicz, 2003) have been used.

6.3.3.3.1 Physics of WF Modulation

Field-effect transistors (Appendix C) are miniature "cousins" of the Kelvin probe. The most common is the insulated gate field-effect transistor. The heart of the insulated gate field-effect transistor is the Metal–Insulator–Semiconductor (MIS) capacitor. Let us form this capacitor from palladium (to be modulated by hydrogen), silicon dioxide (insulator), and p-type silicon (semiconductor), and examine the energy levels in this structure (Fig. 6.32).

In a Thought Experiment, the junction is disassembled (Fig. 6.32) by division through the insulator and the two halves are first treated as electrically isolated objects. In the ensuing equations, we use the common symbol ϕ for the work function of a material. There are three electron work functions to be considered: that of palladium ϕ_{Pd}, that of an arbitrary metal which does not interact with hydrogen ϕ_M, and that of silicon ϕ_{Si}. The insulator is considered to be "ideal" which means that it does not contain mobile charges. Therefore, it does not have a defined Fermi level. Because the two halves are not connected, their energy levels are in an arbitrary undefined position with respect to each other. On the other hand, metal M and palladium (as well as the M and silicon) form ohmic junctions, meaning that the
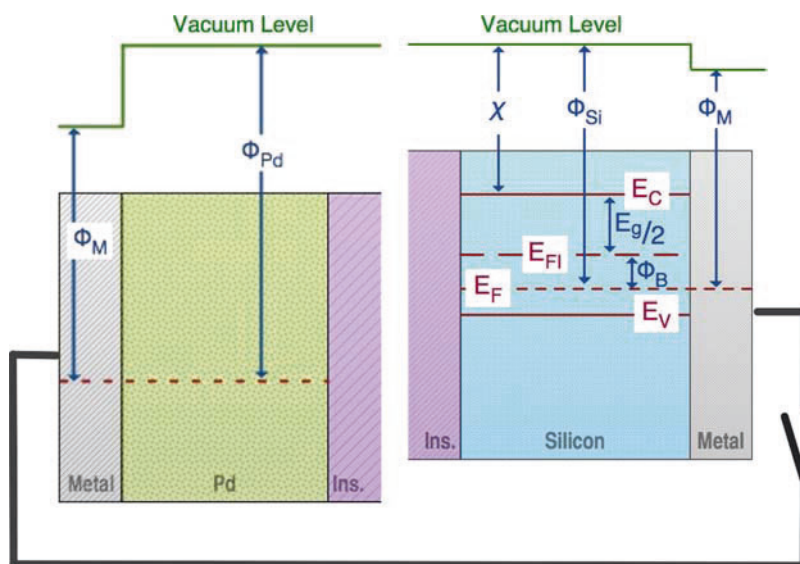
Fig. 6.32 Energy-level diagrams of disassembled palladium gate insulator silicon junction (please see text for discussion)

Fermi levels in those materials must be equal and the contact potentials appear at their interfaces.

In Fig. 6.32, we show graphically both the surface and the bulk terms ($\phi_B + E_g/2$) for the work function of silicon. Let us specify that there is no electric field and therefore no space charge at the insulator/silicon interface. In other words, the potential difference between the surface and the bulk of the semiconductor is zero ($\phi_{SB} = 0$) and the energy bands are flat. This arbitrary, but convenient, state is called the flatband condition. We maintain that condition throughout the ensuing manipulations, with the help of an externally applied flatband voltage V_{FB} if necessary. The Fermi level inside the palladium layer is flat by virtue of the high conductivity (no electric field can exist inside a metal). It is further assumed that no trapped charges or oriented dipoles exist inside the insulator or at its interfaces.

Now we rejoin the two halves (Fig. 6.33). The connecting lead is made again from the same metal M, so that no additional contact potentials develop (they would not affect the final result anyway). We now use the Fermi level in silicon as the starting point and pass the test charge through this structure in a counterclockwise direction until it is returned to the starting point. Work is added at each step: positive when we move up and negative when we move down. The position of the vacuum level above silicon is defined by ϕ_{Si}. Because of the stipulated flatband condition, the vacuum level over the insulator is flat; that is, no electric field is present in the insulator. That defines the position of the Fermi level inside the palladium layer which is ϕ_{Pd} below the vacuum level. Because there is an ohmic contact (exchange of electrons) between M and Pd the Fermi level inside M is the same as that in the Pd layer and defines the position of the vacuum level for M (ϕ_M above the Fermi level). The metal M on the left-hand side of Fig. 6.33 is the same as on the right,
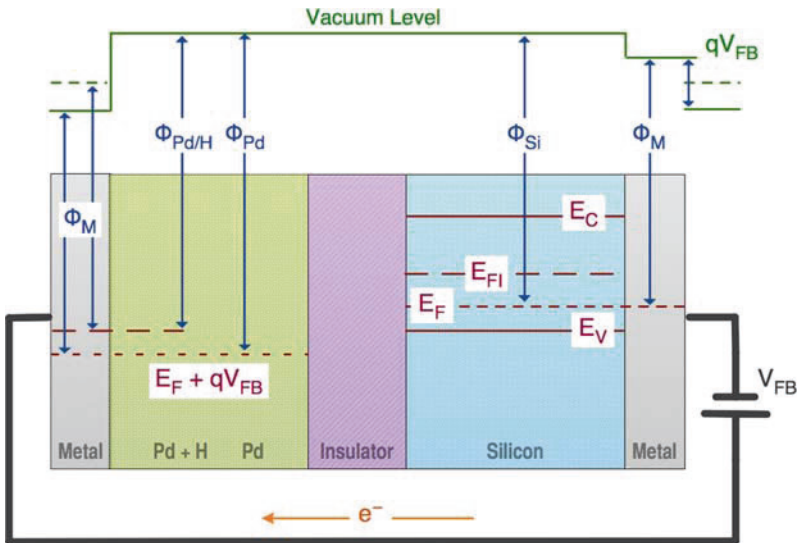
Fig. 6.33 Assembled Pd/insulator/silicon junction

and so is the vacuum level. Because the Fermi levels in M and in Si are equalized, there is a corresponding contact potential at that interface. The sum of the energy contributions in this cycle is then

$$\phi_{Si} - \phi_{Pd} + \phi_M + eV_{FB} + (\phi_{Si} - \phi_M) - \phi_{Si} = 0 \qquad (6.82)$$

or

$$qV_{FB} = \phi_{Pd} - \phi_{Si} = \Delta\phi_{Pd/Si} \qquad (6.83)$$

Now we see that the vacuum levels above the same metal M are not equal. The resulting difference is the flatband voltage V_{FB} which must be externally applied in order to maintain the flatband condition. Thus, the flatband voltage (multiplied by the test charge) equals the difference in the electron work function of Pd and Si. In the nonideal junctions there are other charges and dipoles in this structure that must be added to the overall V_{FB}.

When hydrogen is introduced into palladium at high concentration, the work function of Pd decreases as is indicated by the dashed line in Fig. 6.33. Because the other metal M does not interact with hydrogen (by definition), its work function is not affected, but its vacuum level is displaced upwards by the contact potential. In order to satisfy the flatband condition, V_{FB} must be accordingly reduced, relative to its previous value. Such a change in V_{FB} can be measured. This step is equivalent to adjusting the compensating voltage in the Kelvin probe, as discussed above.

Let us pause and take an inventory of the situation up to this point. (1) We have a plausible mechanism of modulation of both components of WF of a selective layer (palladium) and (2) we have at least two methods of measurement of this effect, the macroscopic Kelvin probe and a field-effect transistor. However, the placement of the selective layer within the structure used for either measurement determines whether the effect is observable. In order to explain this caveat, we add another layer of the same metal M between Pd and the insulator in the structure shown in Fig. 6.33. This would correspond to the "real life" situation when we would try to connect a selective layer by a wire to the IGFET or a Kelvin Probe. It is not necessary to perform the same cycle as we did in Fig. 6.33. Instead, we add the individual energy contributions in the cycle, which begins and ends at the silicon Fermi level (moving again anticlockwise):

$$\phi_{Si} - \phi_M + \phi_{Pd} - (\phi_{Pd} + \phi_M) + eV_{FB} + (\phi_{Si} - \phi_M) - \phi_{Si} = 0 \qquad (6.84)$$

or

$$eV_{FB} = \phi_M - \phi_{Si} \qquad (6.85)$$

We see that the work function of palladium, ϕ_{Pd}, has disappeared from the flatband voltage, the same way as ϕ_M has disappeared from (6.83). This means that, despite the modulation of the work function of Pd (or any other selective layer), in this structure there is no effect on the flatband voltage. The contact potential between Pd and M has changed, but because there are two ohmic junctions between Pd and M, the change of those potentials cancels out. This situation can be generalized by the following rule. In order to observe the flatband voltage change due to chemical
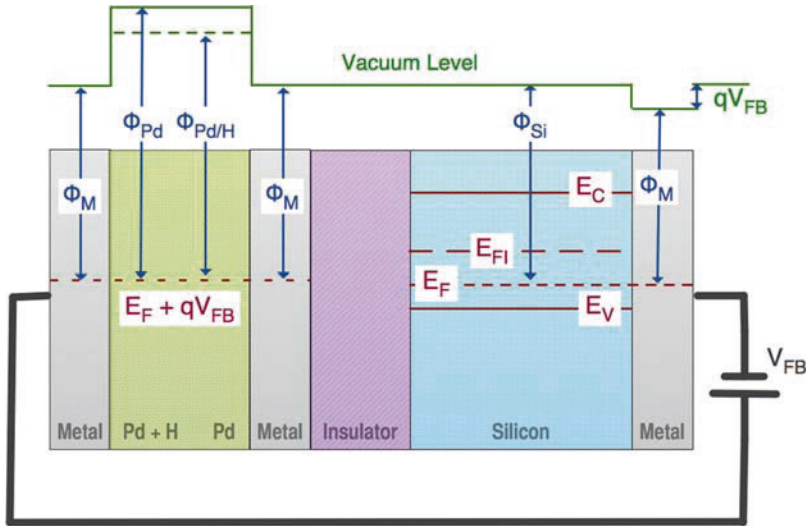
Fig. 6.34 Pd/M/insulator/silicon junction

modulation of the electron work function of the layer, one interface of this layer must be capacitive. This statement can be rephrased more concisely as: *measurable response to the modulation of work function originates only in the layer adjacent to the insulator*. This condition is satisfied in the structure shown in Fig. 6.33, but not in Fig. 6.34.

This condition is naturally also satisfied in the Kelvin probe measurement, where the selective layer is interfaced to the air-gap which is an insulator. However, now we see a subtle but important difference between the Kelvin probe and a MIS-type measurement: in the Kelvin probe, it is the bulk and the surface of the selective layer that contributes the WF modulation and the E_{cell}. In the MIS capacitor (such as in the IGFET), it is the bulk and interface between the solid insulator (e.g., silicon dioxide) and the selective layer that contributes to the overall signal.

The conductivity of the channel region between the drain and source electrodes is modulated by the electric field in the MIS gate capacitor. Thus, the current in the channel becomes a measure of the charge on the silicon "plate" of this capacitor. At this point, we should see the relationship between this device and the Kelvin probe. In the latter, we would have to vibrate the plates in order to nullify the electric field. In the IGFET, we measure the drain-to-source current I_D instead. Another important difference is that in the KP (or the SGFET), the dipoles at both surfaces of the gap contribute to the electric field. On the other hand, the surface dipoles in WF IGFET do not directly contribute to the signal. Only the bulk and the interface are contributing.

The most important point is that the intrinsic Fermi level of silicon, ϕ_{FI}, is not affected by the environment, because the bulk of silicon is well protected by the silicon dioxide and silicon nitride layers and provides the stable and reproducible

reference potential. The only variable term in (6.84) is V_{FB}, which responds to the work function changes of Pd according to (6.83). Therefore, potentiometric gas sensors based on measurement of changes of work function of the selective layer offer this rather unique possibility: *no need for an external reference electrode*. The practical advantage to this consequence lies in miniaturization of gas sensors and in fabrication of gas-sensing arrays.

6.3.3.3.2 Chemical Modulation of WF

Sensing of hydrogen is almost as ubiquitous and important as sensing of hydrogen ions by pH electrodes. Because this reaction is used throughout this chapter, it is necessary to look at its chemistry in somewhat greater detail. The interaction of hydrogen with palladium is highly selective. The mechanism involves several steps in which the molecular hydrogen first dissociates at the Pd surface, from where it diffuses into the bulk of the Pd layer as atomic hydrogen and in doing so forms palladium hydride (PdH_x) over a broad range of concentrations.

$$H_2 + Pd \overset{k}{\Rightarrow} 2H_{ads}$$
$$H_{ads} \Leftrightarrow (PdH_x)_{bulk}$$
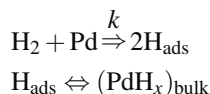
It has been suggested that monoatomic hydrogen adsorbs at the Pd/SiO_2 interface and forms a layer of electrical dipole there (Fig. 6.35a). Thus, at low partial pressure the primary source of the signal is the interfacial dipole whereas at higher concentrations bulk hydride dominates the response. The Pd IGFET operates at elevated temperatures ($\sim$150°C) where the dissociation/diffusion rates for hydrogen are relatively fast. The response in air is shown in Fig. 6.35b. Absorption of hydrogen in Pd is accompanied by volume expansion of the metal lattice (swelling; Butler, 1984). At approximately 3% partial pressure, PdH_x undergoes an "$\alpha - \beta$" phase transition, which may cause cracking of the solid layer of Pd.

It is interesting to note that other species which can be catalytically cleaved to yield hydrogen can also be detected by the Pd IGFET. Ammonia, lower hydrocarbons, and hydrogen sulfide have all been found to produce a signal (Lundstrom and Svensson, 1985).

The general mechanism seems to be the catalytic abstraction of hydrogen from the hydrogen-containing molecules.

$$XH_n + Pd \overset{k}{\Rightarrow} XH_{n-1} + PdH_x$$
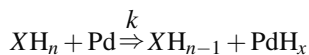
Thus, the Pd layer serves multiple purposes: its surface catalyzes the dissociation of molecular hydrogen, it selectively forms palladium hydride, and it can be used as the metal gate of the field-effect devices. The scheme in Fig. 6.34 also shows the catalytic reaction involving oxygen. If both oxygen and hydrogen are present, the steady-state response of the Pd IGFET includes the surface-catalyzed oxidation.
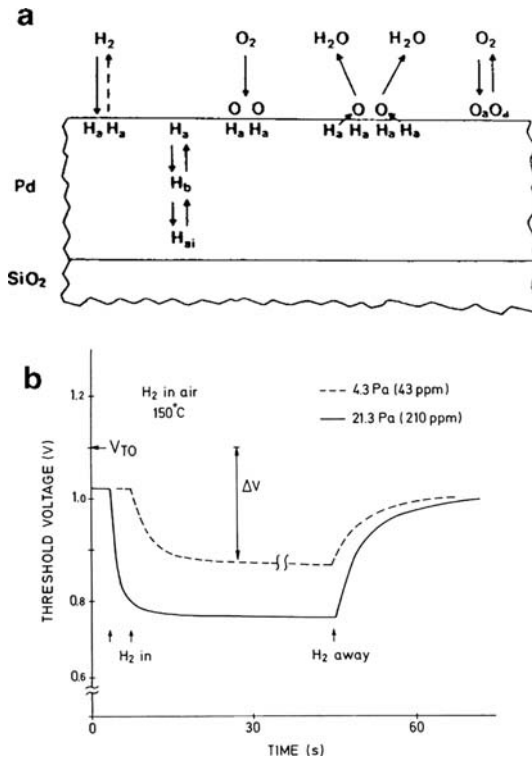
Fig. 6.35 (**a**) Mechanism of hydrogen reaction with Pd gate (adapted from Lundstrom, 1955); (**b**) response of Pd-MOSFET to hydrogen (adapted from Janata and Huber, 1980)

In summary, the solubility of atomic hydrogen in Pd is highly selective, but the catalytic reactions at the surface may interfere with the dissociation process, either by adding or removing H_{ads} to or from the overall scheme. In that respect, oxygen and hydrogen-producing species are the interferants.

As important as it is, the hydrogen–palladium system is rather limited in its application. For this reason, other electronic materials that could operate on the same principle have been investigated. Organic semiconductors proved to be most useful in that respect.

The principle of the operation of work function solid-state sensors is best explained by comparing it with that of equivalent potentiometric sensors for ions. When ions are partitioned between a solution of electrolyte and an ion-selective membrane, an interfacial potential is established that can be viewed as an analog of a contact potential, the contact being made by the transferring ions. This type of charge transfer is described by the Nernst equation, where the interfacial potential is defined per mole. It is important to realize that in both cases an integer value of charge z (i.e., one electron or one ion) is transferred, which alters the electrostatic equilibrium at the interface and leads to an integral value of the slope of the interfacial potential's dependence on the logarithm of the charged species' activity

in the two phases (6.6). In the case of a multivalent ion with charge z, this slope has a value of RT/zF. For example, at 25°C, the slope of the response of an ion-selective electrode is 59 mV/decade of change in concentration of the univalent ion, 29 mV/decade of concentration change of the divalent ion, and so on.

The situation is substantially different in the case of chemical modulation of the work function (Janata, 1991). When electrically neutral molecular hydrogen dissociates at the surface and dissolves in palladium, it partially donates charge density to the free electrons in the metal and changes the position of the Fermi level and thus its work function. Consequently, the contact potential between palladium and copper changes as well. The origin of the contact potential is again the distribution of electrons between Pd and Cu, but the cause of this distribution is the interaction between palladium and hydrogen.

In order to extend the analytical usefulness of this type of interaction, we examine the more general reaction between neutral gas molecules and a semiconductor. Because the Fermi level is defined with respect to one electron, the slope of the work function response is expressed in multiples of kT. If a donor molecule, for example, gas G, enters the semiconductor matrix, it transfers (or accepts) a fraction of its electronic charge to (or from) the semiconductor according to

$$G = G^\delta + \delta e \tag{6.86}$$

This charge-transfer equilibrium can be written as

$$K_G = \frac{[e]^{2\delta}}{SP_G} \tag{6.87}$$

where K_G is the equilibrium constant, P_G is the partial pressure of the gas, and S is the solubility of G in the solid phase. Thus the product αP_G is the concentration of the dissolved gas G in the solid phase.

The ease with which a molecule forms an ion, either by accepting or donating one electron, is given by the Mulliken electronegativity coefficient χ. Numerically, this is the average of its molecular electron affinity E_a and its ionization potential I_p.

$$\chi = 0.5(E_a - I_p) \tag{6.88}$$

On the other hand, the ability of a semiconductor to donate or accept electrons is uniquely related to the energy of the electron in its Fermi level E_F. The transfer of fractional charge δ_e can be viewed as a redox equilibrium between the dopant and the matrix in which the role of the electron donor and of the electron acceptor is relative and governed by the difference between χ and E_F, respectively.

$$\delta_e = \xi(E_F - \chi) \tag{6.89}$$

The proportionality coefficient ξ is discussed later.

The Fermi energy (level) is derived from the Fermi–Dirac statistics which describe the distribution of indistinguishable, noninteracting particles in n available

energy levels. It can be used for electrons or holes, the chief formal difference being in the presence of the degeneracy factor g (which in silicon is 2 for a donor and 4 for an electron acceptor). The values of g for an organic semiconductor are not known a priori. Otherwise, the arguments concerning the interactions of the donors and acceptors are the same. The non-Fermi–Dirac distribution of indistinguishable interacting particles has been also developed (Reiss, 1985). Both forms of the statistics of the particle distribution lead to the same final result in this analysis.

The occupancy of the levels by the donor molecules D is

$$N_{D^+} = N \left[1 - \frac{1}{1 + \frac{1}{g_D} \exp \frac{E_D - E_F}{kT}} \right] \tag{6.90}$$

where $N = N_D + N_{D^+}$, the Fermi level E_F is the average value of energy of electrons in the phase, and E_D is the dopant energy level. The basic premise of calculation of the Fermi level is the condition of charge neutrality in the bulk of the phase. In other words, there is no internal electrical field in the bulk of the semiconductor. As long as the charge remains inside the phase, the Fermi level E_F equals the chemical potential of the electron.

Rearrangement of (6.90) yields

$$E_F = E_D + kT \ln \frac{N_D}{g_D N_{D^+}} \tag{6.91}$$

Thus, E_F varies logarithmically with the concentration of the primary dopant. The energy band structure of organic semiconductors is more complicated than that of silicon. Nevertheless, the coupling of the secondary dopant, such as a gas molecule, is expected to occur through one of the energy levels defined by the primary dopants. For the time being, it is not necessary to specify what this state may be. In the case of some organic semiconductors it could be the polaron or the bipolaron state, which have been shown to form charge-transfer complexes with various organic vapors (Blackwood and Josowicz, 1991; Li et al., 1996).

We now use N as the total primary dopant concentration (e.g., the donor) and link it to the charge transfer from the gas molecules through the electron-exchange equilibrium. The ionization equilibrium involving this level is

$$N_D = N_{D^+} + e \tag{6.92}$$

$$K_D = N_{D^+} \frac{[e]}{N_D} \tag{6.93}$$

Combining (6.87) and (6.92) then yields

$$\frac{N_D}{N_{D^+}} = (N_G S P_G)^{\frac{1}{2\delta}} \tag{6.94}$$

Substitution into (C.6) from Appendix C then yields for an n-type semiconductor

$$E_F = E_D + \frac{kT}{2\delta_e} \ln \frac{K_G \alpha}{(g_D K_D)^{2\delta}} + \frac{kT}{2\delta_e} \ln P_G \tag{6.95}$$

or

$$E_F = E_{D^*} + \frac{kT}{2\delta_e} \ln P_G \tag{6.96}$$

where the standard dopant energy level at unity partial pressure of gas G is

$$E_{D^*} = E_D + \frac{kT}{2\delta_e} \ln \frac{K_G S}{(g_D K_D)^{2\delta}} \tag{6.97}$$

Equation (6.96) is the relationship between the position of the Fermi level in an n-type semiconductor, and the partial pressure of the donor molecules in the gas phase. A similar relationship can be derived for the p-type semiconductor for which the ionization of a discrete acceptor state N_A is given as $N_A + e = N_A-$.

The resemblance of (6.96) to the Nernst equation is more than fortuitous. The fundamental difference lies in the fact that the change of the experimentally observable parameter E_F with the partial pressure of dopant gas has a fractional value given by δ_e, whereas in the Nernst equation the slope has an integral value given by the charge z on the partitioning ion. Thus (6.96) can be viewed as a more general form of the Nernst equation that includes the charge-transfer modulation of the Fermi level of electronic semiconductors. The weak point of this treatment has been the assumption that the enthalpy of the guest–host interaction remains constant over the concentration range but at the same time accepting its modulation. That assumption implies constancy of the Fermi level. Because the position of the Fermi level dominates the interaction that contradiction renders the model only semiquantitative. Nevertheless, acceptable experimental correlations have been obtained (Topart and Josowicz, 1992a, b).

When the transistor in the saturation region (i.e., $V_D > V_G - V_T$) is operated in constant current mode, the gate voltage V_G reflects the changes in the work function of the sensing layers and thus becomes its output. Therefore, from (C.29) we obtain

$$V_G = \left(\frac{2LI_D}{\mu_e C_0 W} \right)^{1/2} + V_T^* \tag{6.98}$$

Combining (6.98), (C.19b), and (C.26) yields dependence of V_G on partial pressure of the donor–acceptor gas. In direct analogy with ISFET operation in the presence of interfering ions (6.71) we can write the following for the transistor in some arbitrary gas phase and the analyte.

$$V_G = V_G^* + \frac{kT}{2\delta_G} \ln \left(P_G + \sum_i K_i P_i \right) \tag{6.99}$$

where V_G^* contains all constant terms related to the transistor and the selective layer, as well as the square root of the drain current. When the transistor is placed in the
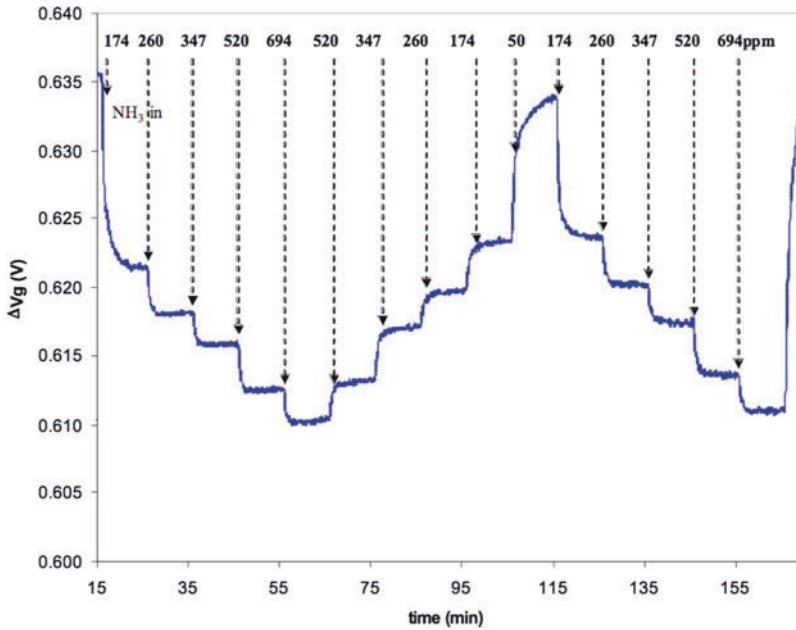
Fig. 6.36 Response of PANI/CSA/IL WF CHEMFET to ammonia (adapted from Saheb, 2008)

"zero gas," in which the partial pressure of the analyte gas G is zero ($P_G = 0$), the output is

$$V_G^0 = V_G^* + \frac{kT}{2\delta_G} \ln \sum_i K_i P_i \qquad (6.100)$$

It represents the baseline value of the response. The step changes of the concentration of the analyte elicit corresponding step changes of the output voltage which are obtained by subtracting (6.100) from (6.99).

$$\Delta V_G = \frac{kT}{2\delta_G} \ln \left(\frac{P_G}{\sum_i K_i P} + 1 \right) \qquad (6.101)$$

Implicit in this equation is the assumption of constancy of the selectivity coefficient K_i and the assumption of constancy of the term V_G^*. Those assumptions put this otherwise useful equation in the category of empirical relationships, the same as the Nikolskij–Eisenman equation. An example of such a response is shown in Fig. 6.36 in which WF CHEMFET with doped polyaniline/camphorsulfonic/ionic liquid gate was exposed to stepwise changes of ammonia concentration. In this case ammonia acts as an electron donor, thus lowering the work function of the selective layer. The value of δ_G determined from the slope of (6.101) was found to be 0.6.

Formation of charge-transfer complexes is a common occurrence in organic chemistry; there are also many analytical reactions that utilize this type of interaction. The materials particularly suitable for this type of transduction are organic

semiconductors, because they can be penetrated by a wide variety of gases. They are also the materials on which the validity of (6.96) has been experimentally verified. The direct confirmation of the relationship between the electron affinity of the polymer matrix (expressed as the initial value of the WF) and its response to the secondary doping was made. In those experiments, various formulations of conducting polymers were prepared and the value of their initial work function in nitrogen was determined with the Kelvin probe. The polymers were based on polypyrrole and p-polyphenylene doped with different levels of different anions. This provided a series of materials with the WF ranging between $+60$ and $+615$ mV (versus Au reference plate). The polymers were then exposed to a stream of the following vapors (in nitrogen): CH_3OH (4.6 mM); i-C_3H_7OH (0.8 mM); n-C_6H_{14} (1.6 mM); and $CHCl_3$ (1.2 mM). The concentrations were determined by gas chromatography. Following the exposure to a given vapor, WF reached an equilibrium value in less than 60 s. In those experiments, the observed modulation of WF was the difference between the vapor and the nitrogen value for each polymer (Fig. 6.37).

The results in Fig. 6.37 show that a given gas can behave with respect to the matrix either as a donor or as an acceptor of electrons depending on the initial value of the WF of the polymer. Because the gas type and the doping concentration remained constant throughout those experiments, the only variable parameter was the initial work function value (WF_{init}) of the polymer. Thus, the slope of the
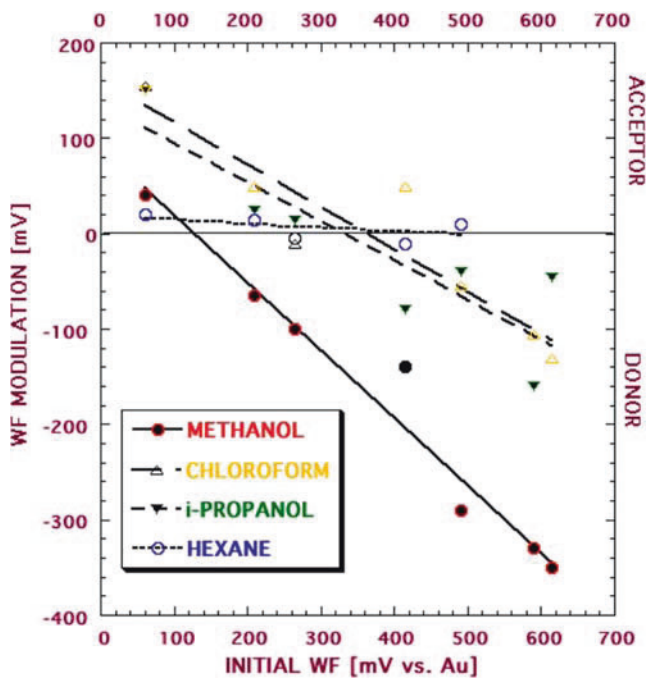
Fig. 6.37 Modulation of initial work function of various polymers and their response to formation of charge-transfer complexes with organic vapors (adapted from Josowicz, 1988)
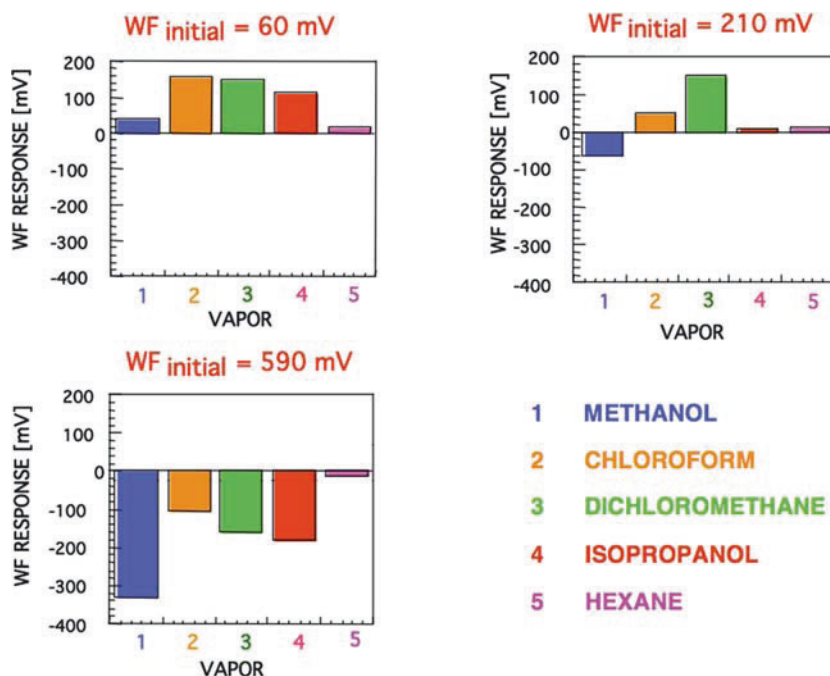
Fig. 6.38 Selectivity diagrams obtained by modulation of the initial work function of the polymer (adapted from Josowicz, 1988)

lines in Fig. 6.37 is the coefficient ξ (6.89). The points at which these lines intercept the "zero" WF modulation lines correspond to the condition when no charge transfer takes place; that is, $\delta_e = 0$ and $E_F = \chi$ for a given gas molecule/polymer combination. Responses to various vapors converted to relative sensitivity diagrams are shown in Fig. 6.38.

It is possible to obtain the value of fractional charge δ from the dependence of Δ WF on partial pressure of the analyte gas (P_G) (6.96). This was done for the modulation of WF of poly(phosphazene)/benzoquinone layers containing iodine by tributyl phosphate (TBP) vapor (Li et al., 1996) and for polypyrrole by methanol (Blackwood and Josowicz, 1991; Topart and Josowicz, 1992a, b). The corresponding reaction, caused by charge transfer between iodine and TBP, is given by the following equilibrium reaction.

The δ_e value for this reaction was determined from the slope of the Δ WF versus log C_{TBP} plot as 0.82 (Fig. 6.39).

It should be noted that at the low concentrations (ppb range) the observed change of WF is apparently dominated by the surface contribution. For this determination, it is also necessary to know the solubilities α of the vapors in the polymer. They can be determined gravimetrically as was done for similar measurements for methanol. Inasmuch as the charge-transfer complex is a dipole, it is possible that the solubilities of these vapors in the polymer are some function of the E_F.
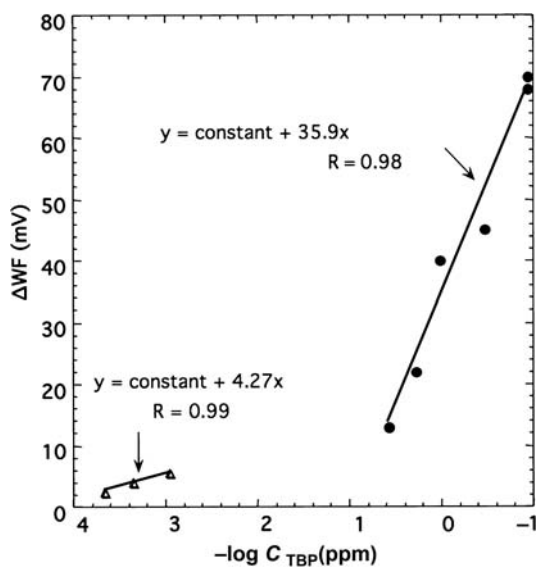
Fig. 6.39 Response of WF-FET with poly(phosphazene/benzoquionone) gate to low and high concentration of tributyl phosphate (adapted from Li et al., 1996)

In this derivation, it was assumed that the charge-transfer complex is formed between the p- or n- primary dopant and the gas acts as a secondary dopant. In fact, the interaction of the secondary dopant with any energy state in the matrix is possible; that would lead to the same result, as long as the exchanged electron density becomes part of the electron population governed by the Fermi–Dirac statistics. The analytical utility of this relationship has been shown for several inorganic gases (Janata and Josowicz, 2003).

6.3.3.4 High-Temperature Ionic Sensors

A potentiometric electrochemical cell consisting of a reference electrode, solid-state electrolyte(s), and an indicator electrode can provide information about the partial pressure of gas in the same way as the cells utilizing ion-selective electrodes and liquid electrolytes can. The general mechanism is as follows. A sample gas, which is part of a redox couple, permeates into the solid-state structure usually through the porous metal electrode and sets up a reversible potential difference at that interface according to the reaction

$$X_{gas} \pm e \Leftrightarrow A^{\mp} \tag{6.102}$$

Three requirements must be met in order to obtain a valid measurement. First, we need a stable reference potential at one interface, which is not affected by the changes of the composition of the sample. This is the usual requirement of

a reference potential in any potentiometric measurement. Second, there must be no electronic conduction through the solid electrolyte(s), otherwise the cell would internally short out. Third, in the case of multiple-layer solid electrolytes, there must be a charge-transporting species identified at each interface. The first requirement can be satisfied by providing a constant (reference) partial pressure of gas to one side of the otherwise symmetrical cell. This is the reference gas electrode arrangement, which is analogous to a liquid electrolyte concentration cell. An example of this type of sensor is the high-temperature potentiometric oxygen sensor based on yttrium-stabilized zirconia (YSZ; Fig. 6.40).

The schematic representation of the cell is

$$O_2(\text{ref.}), Pt /\!/\!/ ZrO_2; Y_2O_3 /\!/\!/ Pt, O_2(\text{sample})$$
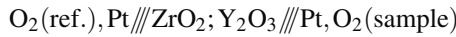
The symbol "///" indicates the porous metal. The electrochemical reactions taking place at the two electrodes are identical.

$$1/2O_2 + e \Leftrightarrow 1/2O_2^-$$
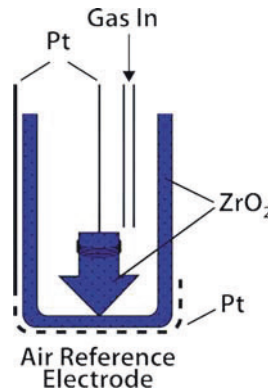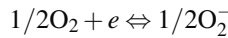
Fig. 6.40 Potentiometric gas sensor based on yttrium-stabilized zirconia (YSZ)
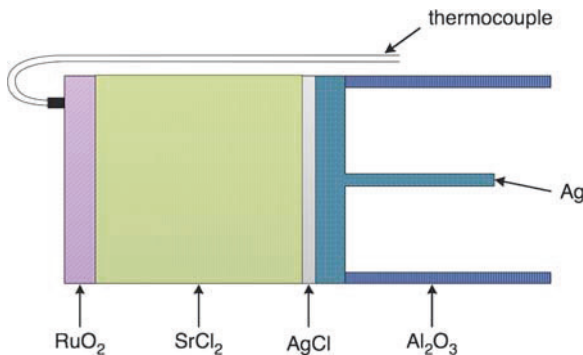
Fig. 6.41 High-temperature chlorine sensor (adapted from Pelloux et al., 1985)

At operating temperature (100–400°C), the oxygen anions have sufficient mobility in the solid electrolyte. The cell voltage E_{cell} is then related to the partial pressure of oxygen by the Nernst equation written for the concentration cell.

$$E_{cell} = \frac{2.3\Re T}{2F} \log \frac{P_{O_2}(\text{sample})}{P_{O_2}(\text{reference})} \tag{6.103}$$

Another opportunity to realize constant activity of the potential determining ion at the reference interface appears when one chooses the solid electrolyte in such a way that the ion of the redox couple is the same as one ion of the major component of the electrolyte. In that case, the change of the activity due to the electrode reaction with the gas can be neglected against the overall constant activity of that ion in the salt. This is the solid-state reference arrangement. An example is the chlorine sensor (Fig. 6.40), in which the reference potential is set up by the constant activity of Cl^- in the solid AgCl electrolyte. This arrangement is equivalent to a reference electrode of the second kind, discussed in Section 6.2.2.1.

The cell can be written as

$$\overset{\longrightarrow}{Ag} \; / \; \overset{\longleftarrow}{AgCl} \; / \; SrCl_2, KCl \; \overset{\downarrow Cl_2}{/\!/\!/} \; RuO_2$$
$$\underset{Ag^+}{} \quad \underset{Cl^-}{} \quad \underset{\leftarrow Cl^-}{}$$

$$\text{reference} \quad /\text{electrolyte}/\text{indicator}$$
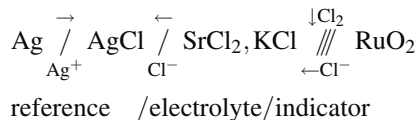
The redox reaction at the indicator RuO_2 electrode (cathode) is

$$Cl_2 + 2e = 2Cl^-$$
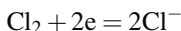
At the reference electrode, the communicating species is Ag^+, and at the AgCl/ $SrCl_2$ KCl interface, it is the chloride ion. Upon increase of chlorine partial pressure in the sample, silver ions are generated at the reference electrode and they combine with the mobile chloride ions in the AgCl layer. The charge balance is maintained by the transport of the chloride ions from the $Sr_2Cl \cdot KCl$ layer. The operation of this sensor is predicated by the low solubility of Cl_2 in the AgCl phase. The operating temperature range of this sensor is 100–400°C. The dynamic range is 1–10^{-6} atm P_{Cl2} and the response time is in minutes.

High-temperature potentiometric sensors for other gases such as O_2, Cl_2, CO/ CO_2, SO_2/SO_3, NO_2, I_2, Na, H_2S, and H_2 have been described (see Table 6.7).

The principle of their operation is the same, but the method of implementation of the sensor is largely dependent on the conditions of the application. Thus, a zirconia sensor for measurement of O_2 in molten steel (1,600°C) has to be designed in such a way that the thermal expansion coefficients of the different layers in this device are matched. On the other hand, a room-temperature potentiometric oxygen sensor can be constructed (Yamazoe et al., 1987) by using another set of materials:

$$\overset{\longleftarrow}{Pt} / \overset{\longrightarrow}{Sn} \; / \; \overset{\longleftarrow}{SnF_2} \; / \; LaF_3 \; / / / \overset{\downarrow O_2}{Pt_{black}} \; \overset{\longleftarrow}{/ Pt}$$
$$\underset{2e}{} \quad \underset{Sn^{2+}}{} \quad \underset{2F^-}{} \quad \underset{OH^-}{} \quad \underset{e}{}$$
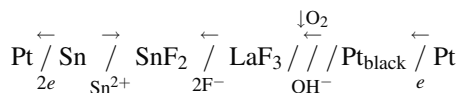
Table 6.7 Potentiometric solid-state gas sensors (adapted from Fouletier, 1982)

Gas	Cell	Temperature range ($^\circ C$)	Pressure range	Accuracy $\triangle P/P(\%)$	Time Lag
O_2	$Ref|ZrO_2 − Y_2O_3|ME, O_2$		O_2-mert $1\,10^{-7}$ atm gas		
$H_2 − H_2O$	Ref Air, $N_1 − N_1O$, Pd–PdO, Co–CoO ME Pt, Ag	500–800	$CO − CO_2$, $H_2−H_2O$, 10^{-8} 10^{-27} atm	2 5	01 s-a fews
$CO − CO_2$					
Cl_2	$Ag|SrCl_2−KCl–AgCl|ME$, Pt, Cl_2 REf $Ag|Ag^+$ Me graphite, RUO_2	100–450	$10^{-6} − 1$ atm	<5	<1mm
SO_2, SO_3	$Ref|K_2SO_4|ME$, $SO_2 + SO_3 + O_2$ $Ag|K_2SO_4−Ag_2SO_4|ME$, $SO_2+SO_3+O_2$ Air, $Pt|ZrO_2 − CaO|K_2SO_4|ME$, SO_3, air Ref $Ag|Ag^+$ ME Pt	700–900	$>10^{-6}$ atm	3	A fews a few mm
	$Ag|Li_2SO_4 − Ag_2SO_4|Pt$, $SO_2 + SO_3$+air	500–750	$10^{-5}10^{-2}$ atm in air		
	Pt, $SO_2 + SO_3 + O_2|Na_2SO_4|Pt$, $SO_2 + SO_3 + O_2$	700	$>10^{-5}$ atm	7	$t_{98} = 8$ mm

Species	Cell				
H_2	$Ref\|HUP\|ME$, H_2 Ref Pd or Pt–H_2, PdH_x	20	$10^{-4}10^{-1}$ atm	15	$t_{-90} \sim 5\,s$
	ME Pd or Pt,moist atmosphere Pd $\|\beta - \beta'' Al_2O_3$ (Na)$\|Pt, N_2 + H_2$			30	1 3 mm
CO	$O_2 + CO$,MR$\|ZrO_2 - Y_2O_3\|ME$, $O_2 + CO$ MR $Al_2O_3 - Pt$ ME Pt	250–350	$0\,5 \times 10^{-4}$ atm in air		t_{90} 5 min
S_x	$Ag\|AgI\|Ag_2S_1$ S (vap)	90–400		50	A few s
	$Ag\|\beta - Al_2O_3(Ag)\|Ag_2S$, S(vap)	90–800			
	$Ref\|CaS-Y_2S_3\|Pt, S_x$	600–900	$<10^{-6}$ atm		
H_2-H_2S	$Ref\|CaF_2\|CaS\|Pt, H_2 + H_2S$	700–950	3×10^{-3} < ratio < 02	5	25 h at 950°C
CO_2	$Ag\|K_2CO_3 - Ag_2SO_4\|Pt$, CO_2 Ref $Ag\|Ag^+$	700–800	$>10^{-6}$ atm		secs to mins
NO_2	$Ag\|B(NO_3)_2 - AgCl\|Pt$, NO_2 Ref $Ag\|Ag^+$	500	$>10^{-6}$ atm		secs to mins
I_2	$Ag\|JAg_4I_5\|Pt, I_2$	40	$>10^{-7}$ atm	3	secs to mins
Na	$NA(vap)\|\beta - Al_2O_3$ (Na)$\|Na$ (vap)	200–360	$10^{-10}10^{-5}$ atm		10 min

In this case, the reversible reduction of oxygen takes place at the LaF_3/Pt (black) interface treated with water vapor. It is known that OH^- and F^- ions can move in LaF_3. Thus, the changes in the concentration of the hydroxyl ions that are part of the redox couple are compensated by the transport of the F^- ions across the Sn, SnF_2/LaF_3 boundary in order to maintain the electroneutrality in the LaF_3 crystal. Oxidation of Sn resulting in the formation of SnF_2 at the Pt/Sn, SnF_2 reference electrode completes the cell reactions. Clearly, the reactions involved in this scheme would not be applicable in aqueous solutions because of the hydrolysis equilibria of the species involved. However, apart from this fact the underlying principles of these sensors and of ion-selective electrodes are essentially the same.

The key knowledge necessary for the rational design and development of high-temperature solid electrolyte sensors is the knowledge of the phase diagrams, solubilities, and the mechanical and electrochemical properties of the materials used in their construction (Weppner, 1987). Because these are equilibrium sensors, the detectable gases are largely limited by the temperature of their operation. The domain of application is inorganic gases whose solubility in the given material of the sensor is usually sharply limited. Therefore, the issue of selectivity with high-temperature potentiometric sensors is less severe than with most other sensors. One major interference that is almost always absent in sensors operating above 100°C is humidity (water vapor).

Food for Thought #6

Adsorption FET

It has been suggested that adsorption of dipolar gas molecules at the surface of the insulator of a field-effect transistor can be used as a principle for detection of gases. It would work in only one type of FET discussed in this book.

1. Which one? Why?
2. Does your answer change if there is a thin layer of gold deposited on the insulator and the gas adsorbs on the outside Au surface?

Contact Potential

3. A glass electrode is connected to the pH meter with copper wire and calibrated with a standard buffer solution. After this calibration, the connecting lead is changed from copper to zinc. Considering the different contact potentials of the Cu and Zn wires, is the calibration still valid or should the electrode be recalibrated?

Liquid Junction Potential

4. Consider an open capillary-type liquid junction. When it is filled with a non-equitransferrant electrolyte (i.e., one with different anion and cation mobilities) a diffusion potential develops that can be measured at the ends of the capillary. Does the value of the diffusion potential change when you change the length of the capillary? Why?

Electrochemical Cell

In a potentiometric measurement with an ion-selective electrode, the cell can be represented as

$$Ag/AgCl/1MKCl, aq.//sample/ISE/1MKCl, aq./AgCl/Ag$$

5. The measurement is done at $i = 0$, yet it is a nonequilibrium system because it contains two "nonequilibrium elements". Identify them and comment on the nature of their "nonequilibrium" state.

Designed "Nonselectivity"

6. There are n types of different ions in solution. A good ISE will have an ionophore that will make it respond selectively only to the ion j. What would happen if you incorporated into the membrane all the ionophores (mobile) for ions $k, l, \ldots, n$? To what would such an electrode respond? Could you take advantage of it in designing an electrode for some highly unique situation? Would this membrane (with many ionophores) be different from the membrane that has no ionophores?

Hydrogel ISFET

7. In a rather unnecessary attempt to "define" the interface between the ion-selective membrane and the dielectric, an "internal hydrogel layer" saturated with, for example, a buffer solution, has been interposed between these two layers. It mirrors the miniaturization of conventional symmetrical membrane ion sensors. The volume of the hydrogel layer is very small. This approach creates more problems than it solves. What is the major problem?

Kelvin Probe

8. The Kelvin probe (a vibrating capacitor) is used to measure the difference in work function of two electronic materials. It is often incorrectly called a Contact Potential Difference (CPD) measurement. Explain why this term is incorrect.

"Perfect" Liquid Junction

9. Leakage of the electrolyte from inside the reference electrode compartment to the sample and diffusion of the sample into the reference electrode are real problems facing the reference electrode. Would it be possible to design a reference electrode in which the open liquid junction channel would be replaced by a Pt wire sealed to the end of the capillary? Such a junction would not leak and the problems would be solved. It could be represented, for example, as

$$Ag/AgCl, aq \cdot KCl(C)/Pt/sample\ldots$$
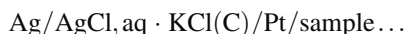
Comment on this idea.

Nonequilibrium Ion-Selective Electrodes

Ion-selective electrodes are designed and routinely used in zero-current potentiometry. It has been suggested that by operating them in a nonequilibrium mode (i.e., by passing a small polarization current through them or by filling them with a concentrated "inner" electrolyte that would diffuse through the membrane) very low detection limits and high "selectivities" could be achieved.

10. Discuss the effect of stirring, adsorption, and change of sample matrix on ISEs operated in such a way.
11. Explain the origin of the potential at the membrane/solution interface.
12. Suggest an experiment in which operation under imposed polarization current could have some beneficial effect.

Reference Electrode

13. An important property of the reference electrode is that it does not change its potential when the sample solution changes. Consider the response as shown in Fig. 1.2. Could you make a reference electrode that would have a very high exchange current density for one ion (i.e., low resistance and no mixed potential), yet would not respond to changes of the activity of such an ion?

Single-Electrode Measurement

The first documented description of an ISFET contained the sensational claim that such a device could be used for the measurement of ion activities without a reference electrode.

14. Discuss the fallacy of that claim from the point of view of performing such a measurement, considering the thermodynamics.
15. Explain the experimental artifact that formed the basis of such a claim.

Symbols

R_{ct}	Charge-transfer resistance
$\Re$	Gas constant ($8.314\,\mathrm{J\,K^{-1}\,mol^{-1}}$)
$\tilde{\mu}_i$	Electrochemical potential
φ	Electrostatic potential (Galvani potential)
a	Activity
t	Time
π	Interfacial potential
π^0	Standard potential
n	The number of electrons
z	The charge of a species
C	Concentration
E_j	Liquid junction (or diffusion) potential
J	Ion flux
D	Diffusion coefficient
E_j	Liquid junction potential
u	Electrolytic mobility
$\mathrm{d}a_i/\mathrm{d}x$	Activity gradient
κ	Conductivity
λ	Ionic conductivity
t_i	Transference number of ion i
Λ	Equivalent conductivity
$\bar{G}_i$	Molar free energy
E_{mix}	Mixed potential
i_a	Anodic exchange currents
i_c	Cathodic exchange currents
K_{sp}	Solubility product
$K_{x,i}$	Selectivity coefficient
K_{ex}	Equilibrium ion-exchange constant
K^{pot}	Potentiometric selectivity coefficient
κ_Δ^{-1}	Thickness of the space charge (the Debye length)
ε	Dielectric constant

V_G	Gate voltage
V_D	Drain voltage
V_T	Threshold voltage
I_D	Drain current
ΔV_G	ISFET response
K_a	Acid dissociation constant
E_F	Fermi level
η	Surface dipole
ϕ	Work functions
V_{FB}	Flatband voltage
K_G	Equilibrium constant of a gas
P_G	Partial pressure of a gas
S	Solubility of a gas
E_a	Molecular electron affinity
δ_e	Fractional charge
ξ	Proportionality coefficient
g	Degeneracy factor
N	Occupancy
E_D	Dopant energy level

References

Allen, F.G. and Gobelli, G.W. (1962) *Phys. Rev*. 127, 150–158.

Arnold, M.A. (1986) *Ion. Sel. El. Rev*. 8, 85.

Bergveld, P. (1970) *IEEE Trans. Biomed. Eng*. BME-19, 70.

Bezegh, K., Petelenz, D., Bezegh, A., and Janata, J. (1987) *Anal. Chem*. 59, 1423.

Bezegh, K., Bezegh, A., Black, G.P., and Janata, J. (1988) *J. Clin. Lab. Anal*. 2, 16–18.

Blackburn, G., Levy, M., and Janata, J. (1983) *Appl. Phys. Lett*. 43, 700–702.

Blackwood, D. and Josowicz, M.J. (1991) *Phys. Chem*. 95, 493–502.

Buck, R.P. (1981) Electrochemistry of ion-selective electrodes. In: J. Zemel and P. Bergveld (Eds.) *Chemically Sensitive Electronic Devices*. Elsevier.

Burr, P.M., Jefferey, P.D., Benjamin, J.D., and Uren, M.J. (1987) *Thin Solid Films* 151, L111.

Butler, M.A. (1984) *Appl. Phys. Lett*. 45, 1007.

Caras, S.D. Petelenz, D., and Janata, J. (1987) *Anal. Chem*. 57, 1920–1923.

Cammann, K. (1985) *Ion-Selective Bulk Membranes as Models for Biomembranes, in Current Topics in Chemistry*, vol. 128. Springer.

Comte, P.A. and Janata, J. (1978) *Anal. Chim. Acta*. 101, 247.

Demoz, A., Verpoorte, E.M.J., and Harrison, D.J. (1995) *J. Electroanal. Chem*. 389, 71.

Dohner, R., Wegmann, D., Morf, W.E., and Simon W. (1986) *Anal. Chem*. 58, 2589.

Domansky, K. and Janata, J. (1995) *Jpn. J. Appl. Phys*. 34(9A), 5054.

Eisenman, G., Bates, R., Mattock, G., and Friedman, S.M. (1966) *Glass Electrode*. Interscience.

Fjeldly, T.A. and Nagy, K. (1985) *Sens. Actuate*. 8, 261.

Fogt, E.J., Untereker, D.F., Norenberg, M.S., and Meyerhoff M.E. (1985) *Anal. Chem*. 57, 1995.

Fouletier, J. (1982) *Sens. Actuat*. 3, 295.

Frant, M.S. and Ross, J.W. Jr. (1966) *Science* 154, 1553.

Franceschetti, D.R., Macdonald, J.R., and Buck, R.P. (1991) *J. Electroanal. Chem*. 138, 1368.

Freiser, H. (1980) Coated wire ion selective electrodes. Chapter 2. In: H. Freiser (Ed.) *Ion-Selective Electrodes in Analytical Chemistry*, vol. 2. Plenum.

Gabrielli, C., Hemery, P., Lettelier, P., Masure, M., Perrot, H., Rahmi, M.I., and Turmine, M. (2004) 570, 275, 291.

Iglehart, M.L., Buck, R.P., and Pungor, E. (1988) *Anal. Chem*. 60, 290.

Janata, J. (1991) *Anal. Chem*. 63, 2546.

Janata, J. (2004) *Electroanalysis* 16, 1831.

Janata, J. and Huber, R.J. (1980) Chemically sensitive field-effect transistors. In: H. Freiser (Ed.) *Ion-Selective Electrodes in Analytical Chemistry*, vol. 2. Plenum.

Janata, J. and Josowicz, M. (2003) *Nat. Mater*. 2(1), 19.

Josowicz, M. and Janata, J. (1988) Suspended gate field effect transistor. In: T. Seiyama (Ed.) *Chemical Sensor Technology*. Elsevier, pp. 153–177.

Lord Kelvin (1898) *Phil. Mag*. 5th series. 46, 82.

Koryta, J. and Stulik, K. (1983) *Ion-Selective Electrodes*. Cambridge University Press.

Kucza, W., Danielewski, M., and Lewenstam, A. (2006) *ElectroChem. Commun*. 8, 416–420.

Lewenstam, A. and Hulanicky, A. (1990) *Ion Sel. Electrode Rev*. 12, 161.

Li, Jing, Janata, J., and Josowicz, M. (1996) *Electroanalysis* 8, 778–783.

Li, Xizhong, Petrovic, S., and Harrison, D.J. (1996) *Anal. Chem*. 68, 1717 and 1726.

Li, Xizhong, Verpoorte, E.M.J., and Harrison, D.J. (1988) *Anal. Chem*. 60, 493.

Liess, M., Chinn, D., Petelenz, D., and Janata, J. (1996) *Thin Solid Films* 286, 252.

Lundstrom, I., Shivaraman, M.S., Svensson, C., and Lundquist, L. (1975) *Appl. Phys. Lett*. 26, 55.

Lundstrom, I. and Svensson, C. (1985) Gas-sensitive metal gate semiconductor devices. In: J. Janata and R.J. Huber (Eds.) *Solid State Chemical Sensors*. Academic.

Maminska, R., Dybko, R., and Wroblewski, W. (2006) *Sens. Actuat. B* 115, 552–557.

Matsuo, T., Esashi, M., and Inuma, K. (1971) Dig. Joint Meet. Tohoku Sect. IEEE J 1.

Moody, G.J. and Thomas, J.D.R. (1978) In: H. Freiser (Ed.) *Ion-Selective Electrodes in Analytical Chemistry*, vol. 1. Plenum.

Morf, W.E. (1981) *The Principles of Ion-Selective Electrodes and of Membrane Transport*. Elsevier.

Moss, S.D., Janata, J., and Johnson, C.C. (1975) *Anal. Chem*. 47, 2238.

Pellat, M.H. (1881) *Journal de Physique* 16, 68.

Pelloux, A., Fabry, P., and Durante, P. (1985) *Sens. Actuat.* 7, 245.

Philips. G. (1951) *J. Sci. Inst*. 28, 342.

Reiss, H.J. (1985) *Phys. Chem*. 89, 3783.

Ruzicka, J. and Lamm, C.G. (1971) *Anal. Chim. Acta*. 54, 1.

Saheb, A., Josowicz, M., and Janata, J. (2008) *Anal. Chem*. 80, 4214.

Sandifer, J.R. (1988) *Anal. Chem*. 60, 1553.

Schenck, J.F. (1977) *J. Coll. Int. Sci*. 61, 569.

Severinghaus, J.W. (1965) In: W.O. Fenn and H. Rahn (Eds.) *Handbook of Physiology*, vol. II. American Physiological Society, Washington, DC.

Smith, R.L. and Scott, D.C. (1986) *IEEE Trans. Biomed. Eng*. BME-33, 83.

Strehlow, H. and Wendt, H. (1961) *Z. Phys. Chem. N.F*. 30, 141.

Topart, P. and Josowicz, M.J. (1992a) *Phys. Chem*. 96, 7824.

Topart, P. and Josowicz, M.J. (1992b) *Phys. Chem*. 96, 8662.

Trasatti, S. and Parsons, R. (1986) *Pure Appl. Chem*. 58, 37.

Vayenas, C.G., Jaksic, M.M., Bebelis, S.I., and Neophytides, S.G. (1996) In: J'O.M. Bockris and B.E. Conway (Eds.) *Modern Aspects of Electrochemistry*, vol. 29. Plenum. pp. 57–202.

Weppner, W. (1987) *Sens. Actuat*. 12, 107.

Williams, R. (1975) *J. Phys. Chem*. 79, 1274.

Yamazoe, N., Hisamoto, J., Miura, N., and Kuwata, S. (1987) *Sens. Actuat*. 12, 415.

Zhang, Tian-Hong, Petelenz, D., and Janata, J. (1993) *Sens. Actuat*. 12, 175.

Chapter 7
Amperometric Sensors

7.1 General Considerations

With this group of electrochemical sensors, information is obtained from the current–concentration relationship. The two most important issues to discuss are (1) the origin of the signal for various types of amperometric sensors and (2) the origins of selectivity. To begin our examination of these issues, we briefly reiterate some of the information presented in the Introduction to Electrochemical Sensors (Chapter 5).

The domain of amperometric sensors from the electrochemical point of view was shown in Fig. 5.1. Depending on whether the electrons are added to or withdrawn from the sample, the indicator electrode **W** is either a cathode or an anode. The general condition of the closed circuit must be again satisfied, but the requirements on the stability of the reference electrode are relaxed as compared to the potentiometric sensors. The chemical transformations that take place at the electrodes on passage of current are governed by Faraday's law (5.3) and by the current–voltage equation (5.11). The form of this equation depends on the geometry of the electrodes and on the experimental arrangement. It is clear that by increasing the difference between the standard and the applied potentials in the exponents of (5.11), it is possible to increase the velocity of any slow electrochemical reaction and to make the charge-transfer resistance small compared to the mass transport resistance. Under those conditions, the current is said to be operating in the "mass transport limited" regime.

Let us consider this regime for the electrochemical reaction between the oxidized (O) and reduced (R) form of a fast redox couple, when both O and R are soluble and only O is initially present in the solution. The redox couple is

$$O + ne \Leftrightarrow R$$

The reduction current is proportional to the area A of the electrode, number of electrons n, and the difference between the bulk concentration C_O^* and the surface concentration $C_O(0,t)$.

$$i = nFm_O A\{C_O^* - C_O(0,t)\} \tag{7.1}$$

The variable m_O is the hitherto unspecified mass transport coefficient of species O, which for the time being is considered constant. As E becomes more negative, the current increases (Fig. 5.1, curve B) until it reaches the plateau limiting current i_L. There it is limited by the mass transport of species O to the surface of the electrode. At this point, every molecule of O that reaches the surface of the electrode is immediately reduced to R. Therefore, the surface concentration of O is zero.

$$\lim_{E \to -\infty} C_O(0,t) = 0 \tag{7.2}$$

This condition holds for both kinetically slow and fast electron transfer reactions, as can be seen from Fig. 5.4. Thus an analytically important relationship is obtained from (7.1), which shows that as long as m_O remains constant, the limiting current i_L is directly proportional to the bulk concentration C_O^*.

$$i_{L,c} = nFm_O A C_O^* \tag{7.3}$$

Throughout this chapter, total cell current i is used instead of the current density j at the working electrode. The two entities are related by (5.15). The reduced form R is produced at the electrode surface. From there, it is transported to the bulk of the solution where its concentration is zero.

$$i = -nFm_R A C_R(0,t) \tag{7.4}$$

Note that the mass transport coefficients m_R and m_O are different and that the concentration gradient is reversed, hence the minus sign. As we have specified above, the electrochemical reaction is very fast, which means that the Nernst equation (5.20b) is satisfied for all values of the surface concentrations of O and R. Thus,

$$E = E^0 + \frac{\Re T}{nF} \ln \frac{C_O(0,t)}{C_R(0,t)} \frac{f_O}{f_R} \tag{7.5}$$

where f_O and f_R are activity coefficients. When (7.1), (7.3), and (7.4) are combined with (7.5), the voltammetric equation is obtained.

$$E = E^0 + \frac{\Re T}{nF} \ln \frac{i_{L,c} - i}{i} + \frac{\Re T}{nF} \ln \frac{m_R}{m_O} \frac{f_O}{f_R} \tag{7.6}$$

The third term is usually close to zero and constant. It can be included in the standard potential, yielding the formal potential E'.

$$E = E' + \frac{\Re T}{nF} \ln \frac{i_{L,c} - i}{i} \tag{7.7}$$

For the current value $i = i_L/2$, the logarithmic term in (7.7) is zero and the half-wave potential $E_{1/2}$ equals the formal potential E. Similar expressions can be derived for

the cases where only R, or both O and R, are initially present in the solution. Thus, for $C_R^* = 0$ and $C_O^* > 0$ (cathodic current), (7.7) applies. Similarly, for $C_O^* = 0$ and $C_R^* > 0$ (anodic current), we can derive (7.8).

$$E = E^{0,\prime} + \frac{\Re T}{nF} \ln \frac{i}{i - i_{L,a}}$$
(7.8)

For $C_R^* > 0$ and $C_O^* > 0$ (cathodic–anodic current) at $i = (i_{L,c} + i_{L,a})/2$:

$$E = E^{0,\prime} + \frac{\Re T}{nF} \ln \frac{i_{L,c} - i}{i - i_{L,a}}$$
(7.9)

Note that subscripts c and a have been added to differentiate the "cathodic" and "anodic" mass transport limiting currents, respectively. These relationships are important in general electrochemistry experiments, because they allow the voltammetric determination of standard potentials.

For sensing purposes, it is necessary to apply constant potential to the working electrode somewhere in the region of the mass transport limiting current plateau, and to follow the changes of the current as a function of concentration according to (7.3). For that relationship to hold, the values of mass transport coefficients m_O and m_R must remain constant during the measurement. These are the conditions of chronoamperometry. As we show, this is true only for very specific conditions. If the solution moves with respect to the electrode, an adherent stagnant layer of solution, called the Prandtl layer, is formed close to the surface of the electrode. In the first approximation, it is assumed that the conditions of linear diffusion apply over this layer. This reasoning is the basis of the Nernst diffusion layer model. This model assumes that the Nernst diffusion layer of thickness δ_O coincides with the Prandtl layer and the mass transport coefficient for species O can be expressed as in (7.10).

$$m_O = \frac{D_O}{\delta_O}$$
(7.10)

D is the diffusion coefficient of species O. The current (7.1) is then

$$i = nFAD_O\{C_O^* - C_O(0,t)\}/\delta_O$$
(7.11)

The thickness of the stagnant layer δ_O is inversely proportional to the tangential velocity of the solution near the electrode. So, as it becomes thinner, the limiting current i_L increases with the stirring rate.

On the other hand, as the Nernst diffusion layer model is applied to an unstirred solution, it is expected that the passage of current will cause formation of the depletion layer (Fig. 7.1), whose thickness δ_O will increase with time. In time, this layer will extend from the electrode surface to the bulk of the solution over tens of μm. In order to estimate the time-dependence of δ_O, we can use the approximate Einstein
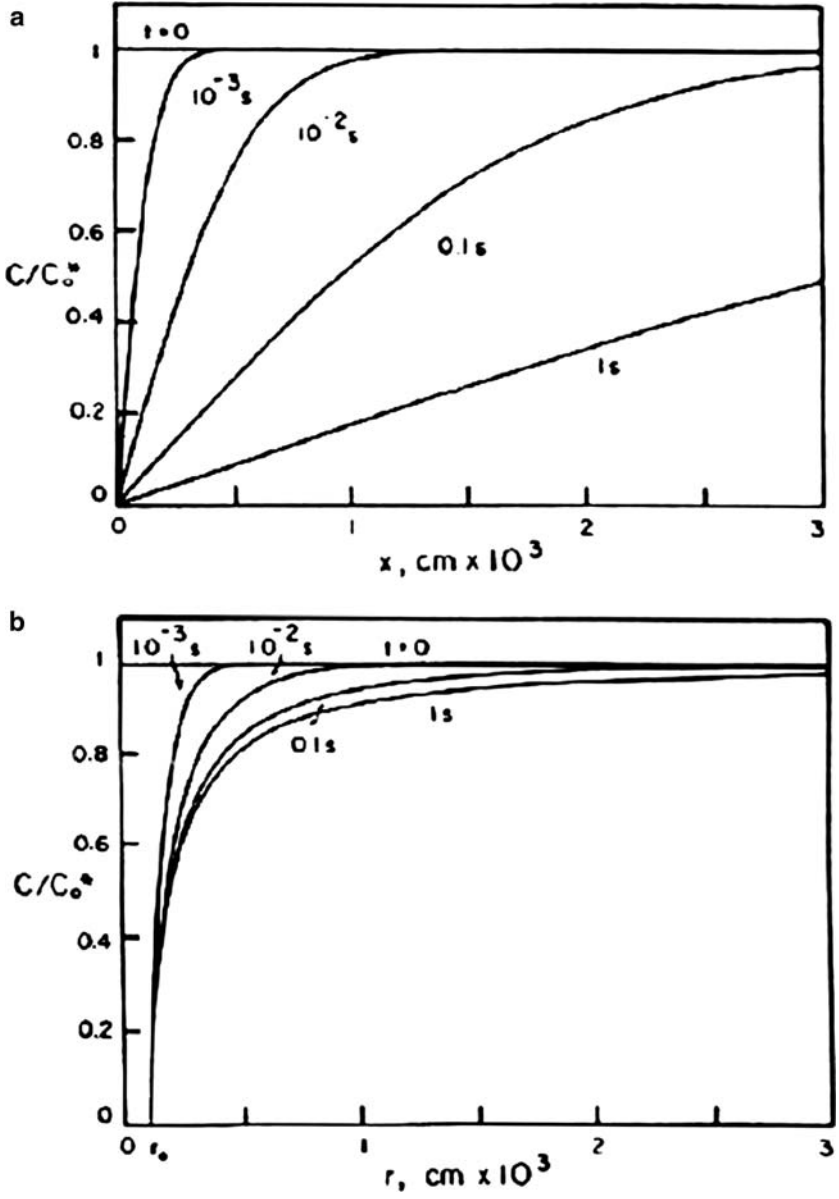
Fig. 7.1 Formation of depletion layer at (**a**) planar microelectrode and (**b**) hemispherical microelectrode (adapted from Wightman and Wipf, 1988)

formula for linear diffusion.

$$\delta_O \approx 2\sqrt{D_O t} \tag{7.12}$$

We may substitute it into (7.10), obtaining a mass transport coefficient as defined by (7.13).

$$m_O \approx \frac{D_O^{1/2}}{2t^{1/2}} \tag{7.13}$$

Therefore, in quiescent (unstirred) solution and under the condition of linear diffusion, the m_O decreases with the square root of time when the potential is applied as a single step to the electrode (7.3). The diffusion limiting current then also decreases with $t^{1/2}$ (Fig. 7.2a).

$$i_L(t) \approx \frac{nFAD_O^{1/2}C_O^*}{2t^{1/2}} \tag{7.14}$$

This equation is only approximate, due to the approximate nature of (7.12). However, it expresses the important characteristics of the experiment done under constant potential, that is, chronoamperometric conditions.

The exact solution for the time-dependence of the current at a planar electrode embedded in an infinitely large planar insulator, the so-called semi-infinite linear diffusion condition, is obtained. Solving the diffusion equation under the proper set of boundary and initial conditions yields the time-dependent concentration profile.

$$C_O(x,t) = C_O^* \left\{ 1 - erf\left[\frac{x}{\sqrt{4D_O t}} \right] \right\} \tag{7.15}$$

where *erf* stands for error function (Bard and Faulkner, 2001).
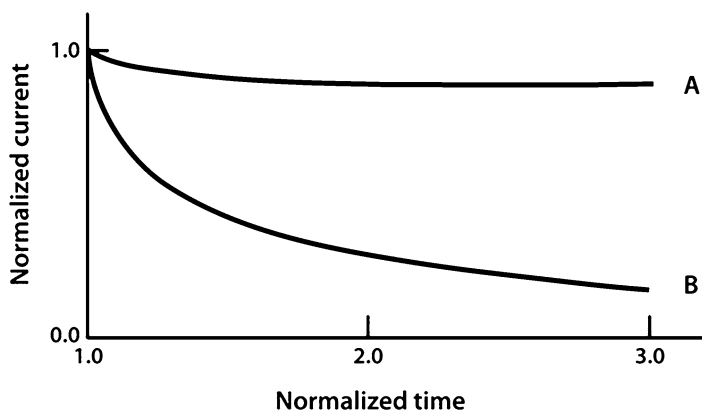
Fig. 7.2 (a) Dependence of current at macroscopic planar electrode placed in unstirred solution; (b) current dependence for hemispherical microelectrode

The derivative of (7.15) with respect to x yields the value of the concentration gradient at the surface of the electrode (at $x = 0$).

$$\frac{\mathrm{d}C_O(0,t)}{\mathrm{d}x} = \frac{C_O^*}{(\pi D_O t)^{1/2}} \tag{7.16}$$

When substituted into Fick's first law (Appendix B, (B.9)), (7.16) becomes the Cottrell equation, which describes the variation of the diffusion limiting current with time exactly,

$$i_L(t) = \frac{nFAD_O^{1/2}C_O^*}{\pi^{1/2}t^{1/2}} \tag{7.17}$$

The concentration profiles for semi-infinite planar diffusion are shown in Fig. 7.1a. The difference between the exact solution (7.17) and the approximate one (7.14) is only 11%.

7.2 Microelectrodes

From the sensing point of view, the time variation of the signal at constant concentration is a big problem. An only slightly better situation exists under the conditions of controlled convective transport. It can be realized by moving the electrode with respect to the solution, for example, by rotating, vibrating, or by placing it in the flowing stream of sample. Such arrangements are generally practical only in sensor systems, but seldom in direct sensing applications. It has been found that for electrodes of certain geometries, with dimensions on the order of a few microns, the current quickly reaches a time-independent, steady-state value (Fig. 7.2b).

$$i_L(t) = \frac{nFAD_O^{1/2}C_O^*}{\pi^{1/2}t^{1/2}} + \frac{nFAD_OC_O^*}{r_0} \tag{7.18}$$

Here, r_0 is the radius of the hemispherical electrode: $A = 4\pi r_0^2$ for a sphere and $A = 2\pi r_0^2$ for a hemisphere. The first term on the right-hand side of (7.18) is the Cottrell term (7.17) and the second is the correction for radial diffusion to the microelectrode. With time, the first term becomes negligible compared to the second. The time t_ε required for the current to reach the steady-state value depends on the desired accuracy ($\varepsilon_\%$) and on the diameter of the electrode $d = 2r_0$ (in μm). It can be estimated by making the first term in (7.18) negligible against the second term, according to the formula

$$t_\varepsilon = \frac{8 \times 10^{-6}d}{\varepsilon_\%^2 D_O} \tag{7.19}$$

Typical values of t_ε for a diffusion coefficient $D = 10^{-5}\,\mathrm{cm^2\,s^{-1}}$ and three common sizes of electrode are shown in Table 7.1.

Table 7.1 Steady-state regime of microelectrodes

Diameter (µm)	t_{99} (ms)	t_{95} (ms)
25	500	200
10	80	32
1	0.8	0.3

Even more important is the shape and the depth of the concentration profile that characterizes the thickness of the depletion layer (Fig. 7.1). The exact solution yields for the concentration profile

$$C_O(x,t) = C_O^* \left\{ 1 - \frac{r_0}{r} erf \left[\frac{r - r_0}{2\sqrt{D_O t}} \right] \right\} \qquad (7.20)$$

The explanation for this behavior comes from the fact that, in a short time, the edge of the depletion layer is close to the electrode surface that looks like a plane. As time progresses, the area through which the molecules enter the depletion layer becomes larger and larger. In other words, the molecules are supplied to the electrode surface from a solid angle whose volume increases with the distance from the electrode and therefore, with the time of electrolysis. The shallow character of the concentration profile at the spherical microelecrodes compared to the deep profile at the planar macroelectrodes is significant from the practical sensing point of view. The sensitivity of the current to the motion of the sample can be seen as the result of changing the thickness of the depletion layer and therefore, the mass transport coefficient. It is much easier to confine the shallow depletion layer of the microelectrode in an additional stagnant layer placed at the surface of the microelectrode than at the macroelectrode. Thus, the mass transport coefficient becomes effectively defined by the thickness of the "stagnant layer" and the response of the sensor is largely undisturbed by the motion of the solution.

The geometry of the microelectrodes is critically important not only from the point of view of the mathematical treatment, but also their performance. Thus, the diffusion equations for spherical microelectrodes can be solved exactly because the radial coordinates for this electrode can be reduced to the point at $r = 0$. On the other hand, a microelectrode with any other geometry does not have a closed mathematical solution. It would be advantageous if a microdisc electrode, which is easier to fabricate, would behave identically to a microsphere electrode. This is not so, because the center of the disc is less accessible to the diffusing electroactive species than its periphery. As a result, the current density at this electrode is nonuniform.

There are other advantages of microelectrodes as compared to the macroelectrodes. Because the current is small, on the order of nA to pA, the voltage drop due to the electrical resistance of the medium is negligible. For this reason, it is possible to perform electrochemical experiments in media that would be otherwise unsuitable for macroelectrodes, such as resistive hydrocarbon solvents and solid electrolytes, without a potentiostat. Secondly, the double-layer capacitance is very small because the area of the electrode is small. This means that very fast modulation experiments

can be performed without paying the penalty of the large charging current transient. Thus, with a 30-μ radius disc electrode, experiments have been performed in the nanosecond range. For the same reason, the signal-to-noise ratio in experiments in which the applied voltage is modulated is much higher. The small physical size of these electrodes is, of course, an important advantage, also for measurements in microliter size samples and in confined spaces, such as those found in, for example, in vivo biomedical applications.

Experiments resulting in steady-state currents at band, cylindrical, and ring electrodes, as well as at electrode arrays, have been described and can be found in electrochemistry textbooks (Bard and Faulkner, 2001). The differences in behavior of these electrodes are significant and should be considered in the design of special sensor applications. The common feature of all these geometries is that the dimensions of the Nernst diffusion layer are comparable with the smallest dimension of the electrode. This is particularly important in microelectrode arrays where the closely spaced electrodes, at steady-state, will behave as a single large electrode whose active area is given by the sum of the areas of the microelectrodes and of the insulator in between. Obviously, this kind of behavior is observed only at times when the Nernst diffusion layers from individual electrodes begin to overlap. The discussion of these effects is beyond the scope of this book and the reader is referred to the specialized literature for more details (Wightman and Wipf, 1988; Fleischmann et al., 1987).

7.3 Oxygen Electrodes

General aspects of amperometric sensing are illustrated on one of the most ubiquitous and important species, oxygen. Amperometric sensors for oxygen come in many different formats. Their applications extend over many fields: from biology and medicine to industrial applications, energy production, and safety. Likewise, the conditions under which the oxygen sensors have to operate range from high temperature (400–1600°C) sensing in gas phase to liquid media sensing in the human body. There is extensive literature devoted specifically to oxygen sensing (Kreuzer et al., 1980; Fatt, 1976; Hitchman, 1978). The principles of operation, the design rules, the problems, and their solutions apply more or less to any electroactive species that can be reduced or oxidized.

At a mercury electrode, in neutral or acidic medium, the reduction of oxygen proceeds in two steps:

$$O_2 + 2H_2O + 2e \overset{E_1}{\Rightarrow} H_2O_2 + 2OH^- \, (E_1 = 0.68)$$

and

$$H_2O_2 + 2e \overset{E_2}{\Rightarrow} 2OH^- \, (E_2 = 1.77)$$
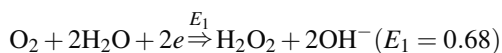
Both reactions are irreversible; therefore, the standard potentials are only calculated. Because E_1 is more positive than E_2, two separate reduction waves should be seen.
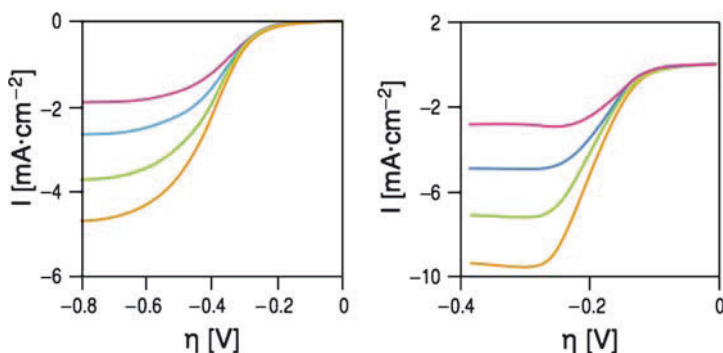
Fig. 7.3 Dependence of the current density on the overvoltage for electrodes rotating at 500 rpm (top curve) and 1,000, 2,000, and 4,000 rpm (bottom curve) under 1 atm O_2 at 37°C. Scan rate: 5 mV s^{-1}. *Left*: polished Pt in 0.5 MH_2SO_4. *Right*: "wired" billirubin oxidase (BOD) on polished vitreous carbon in a pH 7.2 saline (0.15 M NaCl) phosphate (0.02 M) buffer. The "wired" BOD film, loaded at 0.17 mg cm^{-2}, comprised 32.3 wt% BOD, 60.2 wt% redox polymer, and 7.5 wt% PEGDGE (adapted from Mano, 2003)

In fact, this is the case only at the mercury cathode (Koryta et al., 1993). At other electrodes (e.g., Pt, Au, or C) only one four-electron reduction is obtained, meaning that the hydrogen peroxide intermediate is reduced at approximately the same potential as molecular oxygen.

$$O_2 + 2H_2O + 4e \overset{E_1}{\Rightarrow} 4OH^-$$
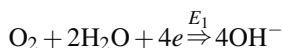
The importance of oxygen reduction extends from chemical sensing to fuel cells and biochemistry. It is one of the most studied electrochemical reactions, yet the fine details of its mechanism are still unclear. There are multiple intermediates involved. One reason for this complexity is the catalytic nature of some of the intermediate steps. This is why the electrode materials and conditions have such an influence on the overall mechanism. The modification of the electrode surface can dramatically change both the electrode kinetics and the mass transport. The example is seen in Fig. 7.3, where reduction voltammograms at a bright Pt electrode and at an electrode modified with biocatalyst are compared (Mano et al., 2003; Soukharev et al., 2004). In this case, the reduction of oxygen does not appear to proceed through the H_2O_2 intermediate stage.

From the sensing point of view, this example serves to illuminate several important points regarding the selection of the optimum operating parameters. They depend on the conditions of the sample, for example, on the presence of interfering species and on pH. In any case, it is important to select the applied potential in the flat range of the mass transport limited current, at the limiting current plateau, where di/dE$\sim$0. The applied potential should be at least 120 mV greater than the half-wave potential. In the case of the example given in Fig. 7.3, it will be $\sim$–600 mV for the bare Pt electrode and $>$–300 mV at the Pt electrode modified with the biocatalyst. Because the mass transport limited current is "flat" the requirement of the

stability of the reference electrode is more relaxed compared to potentiometric measurements. Also, due to the diffusional barrier the current density is almost four orders of magnitude smaller for the modified electrode, which is only a relatively minor inconvenience.

Once the correct applied potential is selected, the mass transport conditions described in Section 7.1 apply. In that sense, oxygen is not different than any other electroactive species. Bare Pt ring microelectrodes of the band thickness between 100 and 500Å are insensitive to flow (Saito, 1967). The original amperometric oxygen electrodes, used mainly in biomedical applications, are bare noble metal wires. They are called monopolar electrodes because an external anode is required for their operation. It goes without saying that they can be used only in a conducting medium and they suffer from fouling up of the electrode surface, which may alter the mass transport properties and may change the response characteristics. This problem can be minimized or eliminated by coating the electrode with hydrophilic polymers, which allow the transport of ions but restrict the contact of higher molecular weight substances (e.g., proteins) with the electrode surface. In this case, we are using mass transport resistance to achieve selectivity. Thin hydrophilic polymers are usually not mechanically strong which may cause problems in applications where the electrode is subjected to mechanical stress, such as insertion into tissue.

7.4 Clark Electrode

A major advance in the performance of amperometric oxygen sensors has been achieved by placing both the cathode and the anode behind the oxygen-permeable membrane (Fig. 7.4). This sensor is known as the Clark oxygen electrode.

Several different membrane materials have been used, namely Teflon, polyethylene, and silicon rubber among others. It is possible to obtain some degree of selectivity by choosing the material of this membrane according to the conditions of the application. The diffusion through such a structure is more complicated. For radial geometry, the steady-state current is given as

$$i = \frac{4\pi F r_0 D_s S_s P(r_1)}{D_s(r_1 - r_0)/D_m r_1 + S_m r_0 r_1} \tag{7.21}$$

where r_0 is the outer radius of the membrane, D is the diffusion coefficient, and r_1 is the radius of the electrode. Subscripts s and m refer to the solution and the membrane, respectively. The S terms describe the solubility of oxygen in the solution and in the membrane. The oxygen partial pressure at the membrane surface is $P(r_1)$. The concentration profile in a radial direction through the 25 μm diameter electrode is compared with the profile for a bare electrode in Fig. 7.5.

From this comparison we see that the depletion effect caused by the electrode itself is almost completely confined within the membrane. This is particularly important if the electrode is used in the medium which itself consumes (or produces) oxygen. Generally, for electrodes in which the depletion field extends beyond
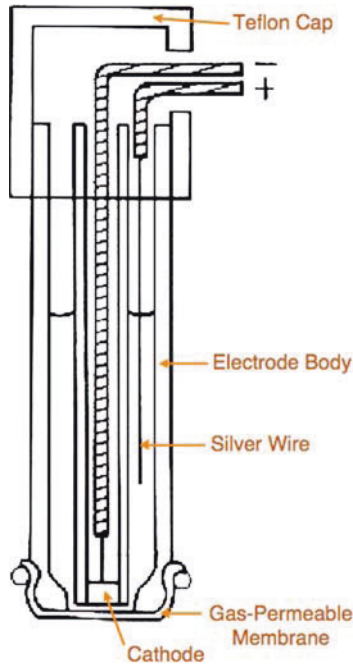
Fig. 7.4 Clark oxygen electrode (adapted from Fatt, 1976)

the membrane, it is not possible to use calibration in aqueous solutions for measurements in the oxygen-consuming tissue. The effect of the depletion field disappears for microelectrodes with a radius smaller than $10\,\mu m$.

The flow sensitivity of the electrode has the same origin, as has been pointed out previously. A stagnant (Prandtl) boundary layer of thickness δ forms around the spherical electrode (radius r_0) placed in the liquid of kinematic viscosity v which is moving with linear velocity U.

$$\delta = \left(\frac{v r_0}{U}\right)^{1/2} \tag{7.22}$$

If the depletion layer is completely inside the stagnant layer, the current is not affected by the change of flow. From (7.18), we know that this happens when the electrode radius becomes small. For Clark-type electrodes, the flow insensitivity is obtained even for larger diameters of the electrode, because of the additional confining effect of the membrane which has lower oxygen transmissivity, $D_m S_m$, than that of the solution.

$$D_m S_m < D_s S_s \tag{7.23}$$

This effect is demonstrated in Fig. 7.6, in which the effect of liquid velocity is compared for several sensors of different dimensions of the membrane and of the electrode radius.
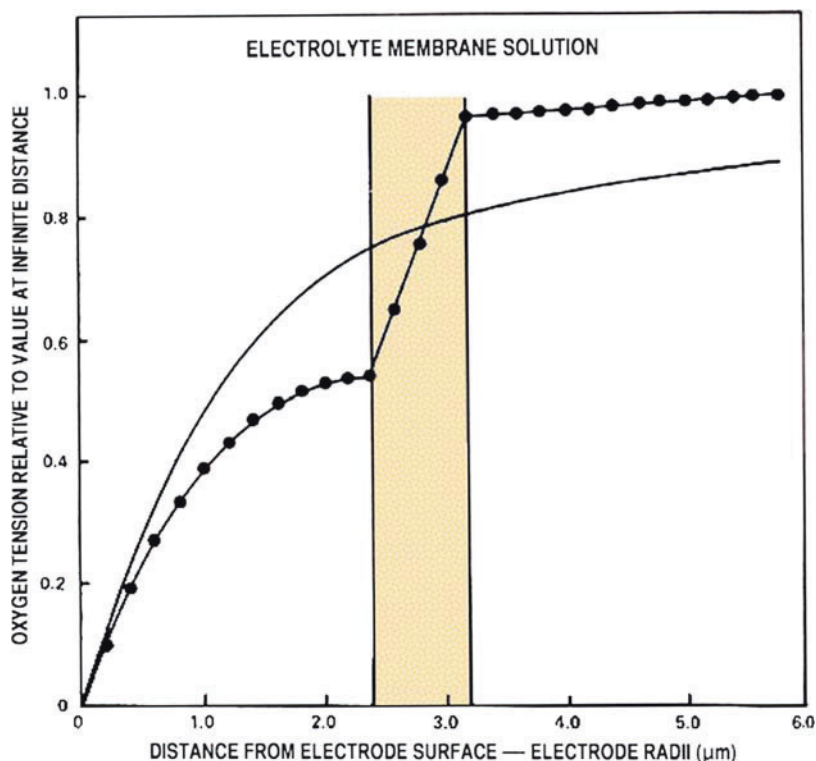
Fig. 7.5 Calculated concentration profiles for oxygen in Clark electrode consisting of (**a**) electrolyte; (**b**) membrane and sample solution. The smooth solid curve is for bare Pt electrode of the same dimensions (adapted from Fatt, 1976)

It would be expected that the speed of response scales again with the radius of the electrode. However, it has been found (Vacek et al., 1986) that the fastest speed of response for a hemicylindrical Clark electrode is obtained with an electrode radius of approximately 5–10 μm. This is due to the fact that as the radius decreases, the effect of the layers that are closer to the electrode surface becomes relatively more important than those that are farther away. In fact, with the further decrease of the radius the time response becomes longer than for the corresponding planar electrodes.

The greatest impact of the Clark oxygen electrodes has been in medicine and physiology. (A schematic diagram of a catheter-size Clark electrode is shown in Fig. 7.7.) On the other hand, a temperature- and pressure-compensated Clark electrode for oceanographic measurements up to 600 ft has also been developed (Fatt, 1976). The normal temperature coefficient of the Clark electrode is ∼2%/°C
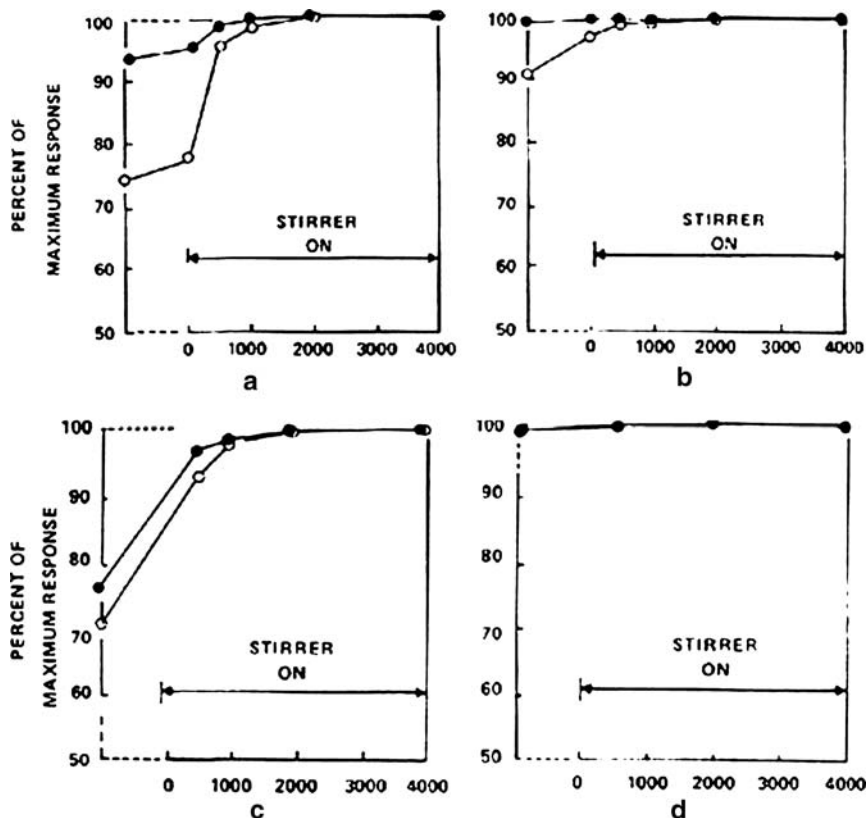
Fig. 7.6 Effect of flow velocity on oxygen electrodes of different dimensions. Open circles represent O_2 – saturated water and full circles represent areated water. The electrodes are (**a**) 50 μm diameter Pt disks; (**b**) 50 μm diameter Pt disks with 25 μm Teflon membrane; (**c**) 300 μm Pt disks; and (**d**) 10 μm Pt disk with 6.25 μm Teflon membrane (adapted from Fatt, 1976)

and the linearity is usually better than 1%. The time response depends mainly on the thickness of the membrane. For a 5 μm thick polypropylene membrane on a 7 μm radius hemispherical electrode, the response time is below 1 s (Hurrell and Abruna, 1988).

The idea of confining the electrical circuit behind a gas-permeable membrane is quite general. The similarity between the potentiometric Severinghaus electrode (Section 6.2.2) for CO_2 and other hydrolyzable gases is obvious. Amperometric electrodes for sensing of other electroactive gases, such as H_2S, NO, NO_2, Cl_2, CO, and so on, have been designed. The principles of operation, the design rules, the problems, and their solutions are more or less the same.
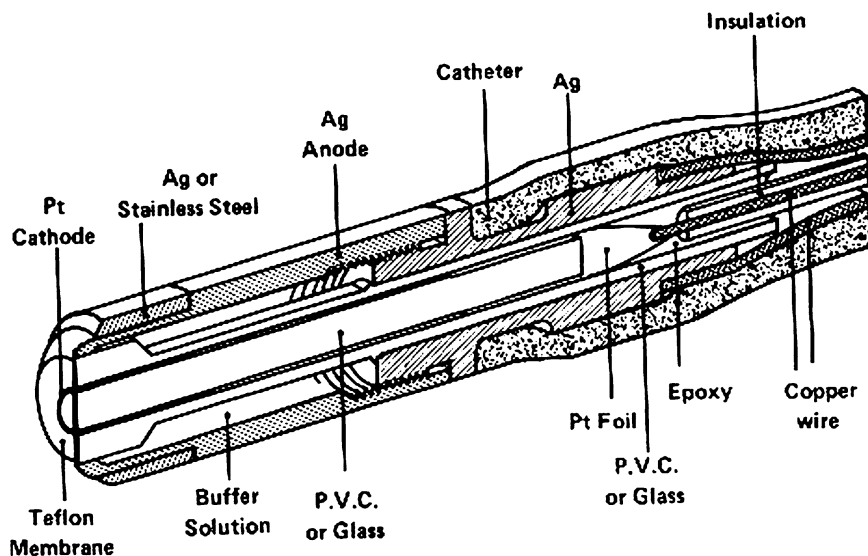
Fig. 7.7 Catheter Clark oxygen electrode

7.5 Amperometric Selectivity

Selectivity can be viewed as the fraction of the overall response of a sensor to the species of interest in the presence of interfering species. For multiple electroactive species detectable by amperometric sensors we can write

$$i_{\text{cell}} = i_1 + i_2 + \ldots \ldots i_R \tag{7.24}$$

where the i terms are the partial currents contributing to the overall cell current i_{cell}. From Ohm's law, each of these partial currents can be expressed as the ratio of one common applied cell voltage E and partial electrode resistance R_n. Because the electrode resistance is the sum of charge-transfer and mass transport resistances, we can write for each species

$$i_{\text{cell}} = \langle E \rangle \left(\frac{1}{R_{\text{ct},1} + R_{\text{mt},1}} + \frac{1}{R_{\text{ct},2} + R_{\text{mt},2}} + \ldots + \frac{1}{R_{\text{ct},n} + R_{\text{mt},n}} \right) \tag{7.25}$$

The equivalent circuit corresponding to this resistive network is shown in Fig. 7.8. The current that carries the information about the analyte flows through the path of the lowest resistance. This seemingly trivial circuit can help us to design the best strategy for selectivity of amperometric sensors.

As we have seen already, there are two kinds of selectivity: thermodynamic selectivity and kinetic selectivity (Chapter 2). Let us first consider how we could use thermodynamic selectivity for amperometric sensors. The electrode and the solution form a double-layer capacitor. The minimum energy of this capacitor occurs at
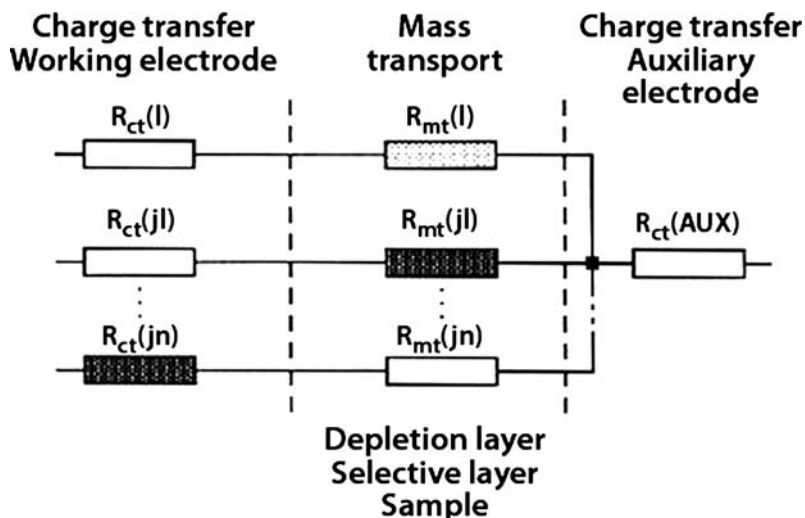
Fig. 7.8 Origins of selectivity of amperometric sensors based on charge-transfer resistance and mass transport resistance

the point of zero charge, when the electric field at the electrode interface is zero. As the applied potential of this capacitor is increased, a point is reached when the capacitor breaks down, electrons are transferred between the solution and the electrode, and electrolysis takes place according to Faraday's law. Let us assume that compounds (Ox_1, Ox_2, ...) in their oxidized form are present in the solution and that they are all electroactive at different applied potentials, which means that their ability to exchange electrons with the electrode is different. Let us also assume that these are fast, electrochemically reversible reactions; that is, their charge-transfer resistances are very low. The more positive their standard potentials, the more easily they will be reduced.

As the potential is scanned from positive to negative, the reduction of Ox_1 takes place first. As the potential is made even more negative, Ox_2 begins to be reduced, then Ox_3, and so on. Thus, at the applied potential E_1, only Ox_1 will be reduced, but at the more negative potential E_2, simultaneous reduction of Ox_1 and Ox_2 will take place. In order to determine these two species separately, measurements at two potentials must be made. In order to do that, the two potentials have to be at least 180 mV apart. Given an electrochemical window of ~2.5V, we can see that the maximum number of electroactive species that can be accommodated is not more than 13, provided that their standard potentials are equally spaced. In reality, the number of different species that can be selectively determined in a mixture by using the selection of the applied potential is 4–6 at most. Thus, the choice of applied potential offers only a very limited selectivity and is used only to complement other modes of selectivity. In the context of the equivalent electrical circuit (Fig. 7.8), this strategy would be represented by the same potential applied to all resistive channels

at once. The same argument can be made about increasing the temperature because, again, all rate constants are affected similarly.

A more fruitful approach to higher selectivity is to increase the difference between the total electrode resistance to the interferants, relative to the resistance belonging to the analyte. The rate constant of the desired charge-transfer reaction can be increased by electrocatalysis. Thus, for example, incorporation of a highly selective oxidation catalyst such as glucose dehydrogenase (or glucose oxidase) in the electrode proper increases the rate of charge transfer between the electrode and glucose. Thus, the enzyme effectively and selectively lowers the charge-transfer resistance. Both inorganic and organic catalysts have been used to selectively lower the charge-transfer resistance of selected substrates.

Manipulation of the mass transfer resistances is another possibility. Let us assume that the analyte is an electrically neutral species, but the major interferant is charged. By placing an ion-exchange membrane with immobile charge of opposite polarity to that of the interferant in front of the electrode, the access of the charged interferant becomes blocked by the electrostatic repulsion. These selectivity design strategies can be summarized in a statement that applies also to other life situations. *In amperometric sensors, the information is obtained from the current path of least resistance.*

7.5.1 Modified Electrodes
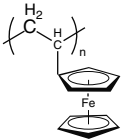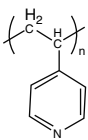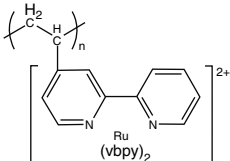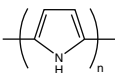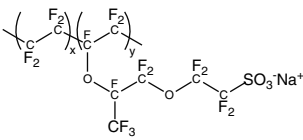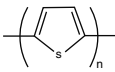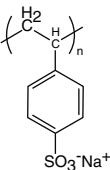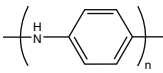
A large amount of research has been done on chemical modification of electrodes. The authoritative treatment of this subject can be found in Bard and Faulkner (2001). Because it is a very active area of electrochemistry, this subject is being periodically reviewed. From the sensing point of view, the motivation for electrode modification has been to introduce additional flexibility in the design of, and additional control over, the electrochemical processes taking place at the electrodes. We have seen one example of such a modification already (Section 7.3; Soukharev et al., 2004).

In general, traditional electrode materials are substituted by "electrode super-structures" designed to facilitate a specific task. Thus, various modifiers have been attached to the electrode that lower the overall activation energy of the electron transfer for specific species, increase or decrease the mass transport, or selectively accumulate the analyte. These approaches are the key issues in the design of chemical selectivity of amperometric sensors. The long-term chemical and functional stability of the electrode, although important for chemical sensors as well, is typically focused on the use of modified electrodes in energy conversion devices. Examples of electroactive modifiers are shown in Table 7.2.

To this table we can also add nonconducting blocking layers that selectively hinder access of certain species, for example, interferants to the electroactive part of the electrode.

Any additional layer that modifies the metal electrode has a certain thickness. That thickness can range from 1–2 nm to thousands of nanometers. This opens up

Table 7.2 Examples of modifiers for amperometric sensors (adapted from Bard and Faulkner, 2001, pp. 584–585)

many possibilities for the interaction of the analyte **X** with such an electrode. Let us assume that the mass transport in the solution is not an issue; that is, it has been taken care of by controlled convection or by making the electrode small enough to operate in the radial diffusion regime. The real question then becomes: where does the electron transfer take place? The schematic depiction of the possible modes of interaction of the analyte is shown in Fig. 7.9. To that model is added one quite common possibility: penetration of **X** to the interior of the layer where it undergoes the appropriate redox reaction. It is again the case of the diffusion–reaction mechanism that was described in Chapter 2 for conversion of substrates in enzyme-containing gels. However, in the case of redox electrodes, the boundary between the electrode and the modifying layer has to be described by a nonzero-flux boundary condition. In addition to thickness, the morphology of the modifying layer plays an important role in determining where the actual charge transfer takes place.

We must remember that with amperometric sensors, the analytical information is obtained from the mass transport limiting current. One important consequence of the current–voltage equation is that one can always apply a potential high enough in order to transfer electrons to or from the electrode to a given species of interest.
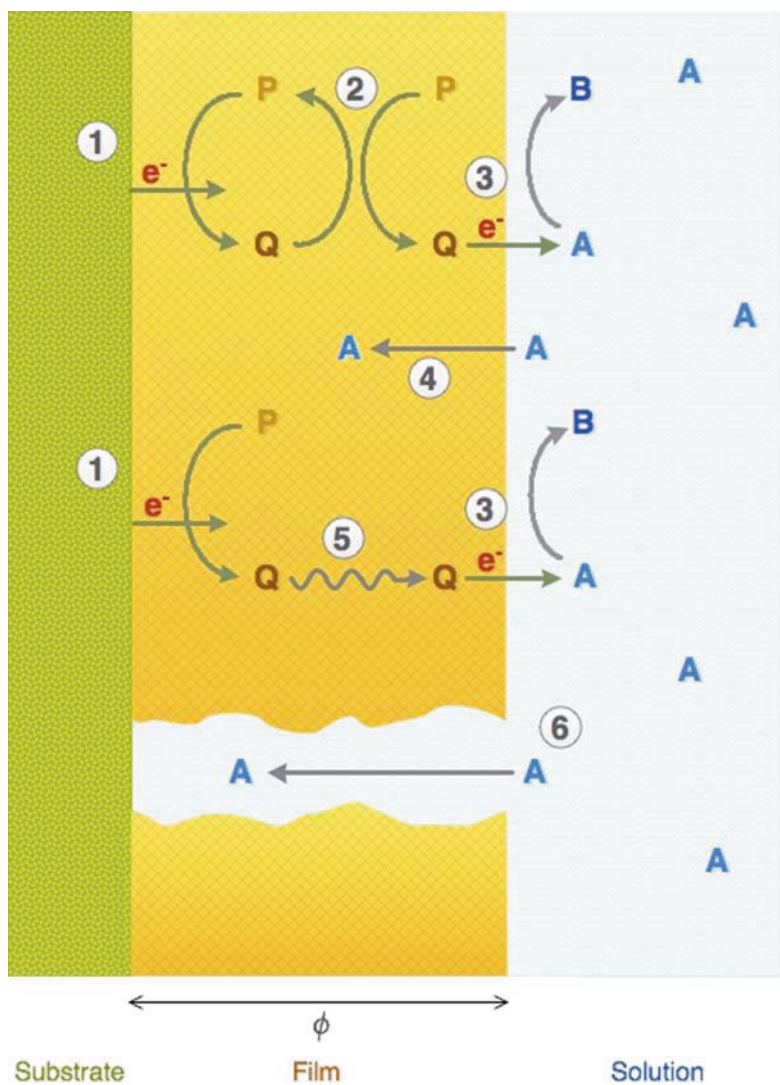
Fig. 7.9 Processes that can occur at a modified electrode. The analyte A penetrates into the film containing reducible agent P. (1) heterogeneous reduction process; (2) successive transfer of electron between reduced molecules Q (5), until the transfer to A at the surface (3); (4) diffusion of A into the film and its reaction with Q; (6) direct penetration of A through the pinhole to the substrate electrode (adapted from Bard and Faulkner, 2001, p. 608)

The selective facilitation of the charge transfer of the species of interest is called electrocatalysis. In such a case, the species of interest are transformed at energies substantially lower than those of the interferants. The higher selectivity therefore implies a lower applied potential at the modified working electrode, which exhibits such selective electrocatalytic properties. In such a situation, the choice of the
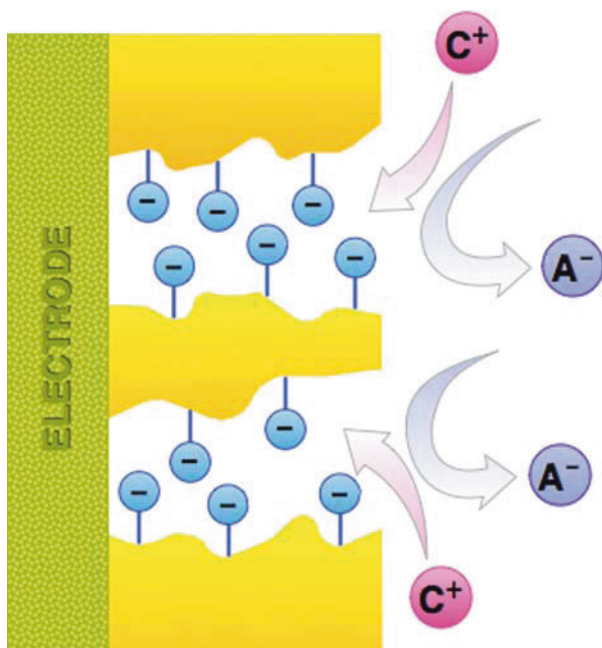
Fig. 7.10 Electrostatic rejection of anions by negatively charged pores in the Nafion membrane

experimental conditions, including the type of modifying layer, is dictated by the presence of specific interferants. The bioamperometric enzyme electrodes discussed in the previous section belong to this category.

It is possible to selectively hinder the access of a species or a group of species to the surface of the electrode. This principle is illustrated on the example of a Pt microelectrode coated with a thin layer of Nafion (Fig. 7.10). Nafion is a sulfonated Teflon. It carries immobile negative charges. It is highly permeable to cations and to the neutral species. Such electrodes have been used successfully for in vivo monitoring [59] of cationic neurotransmitters, such as dopamine, norepinephrine, and 5-hydroxytryptamine in the presence of anionic interferants, ascorbate, 3,4,3-dihydroxyphenylacetic acid, and 5-hydroxyindoleic acid, all of which are negatively charged at physiological pH. These anions are electrostatically excluded from the electrode surface by the negatively charged sulfonic groups present in the Nafion pores. It has been shown that the transport of the interfering anions to the electrode can be inhibited to 0.5–0.7% of the equivalent concentration of dopamine. The oxygen Clark electrode, which has a hydrophobic membrane permeable only to one electroactive species (oxygen), can also be viewed as a perfect example of the "blocking modifying layer."

The third effect that arises from the finite volume of the modifying layer is the possibility of selective and reversible accumulation of the species of interest in the layer (Fig. 7.11) by selective sorption. The analyte is then electrochemically
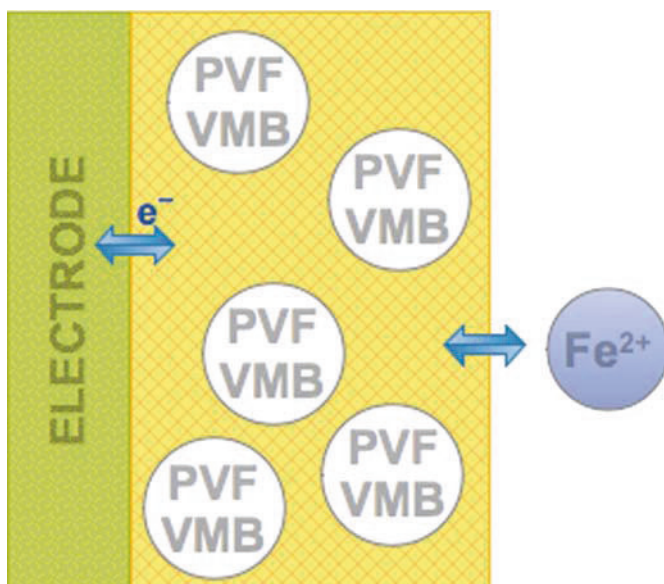
Fig. 7.11 Accumulation of the analyte (Fe^{2+}) in the membrane

transformed, as has been demonstrated with poly (vinylferrocene)/4,3-vinyl-4′-methyl-2,2′-bipyridine copolymer (Hurrell and Abruna, 1988). Ligand 2,2′-bipyridine selectively and reversibly complexes Fe^{2+} ions from the solution. This process is, in effect, an extraction by which the analyte is accumulated at the electrode and then detected. The quantitative relationship between the sample concentration and the concentration at the electrode is governed by the extraction equilibrium, if the layer is thick, or by the adsorption isotherm, if the layer is thin. In either case, the surface-immobilized complex of Fe (II)–(bipyridine)$_3$ can be determined electrochemically.

7.5.2 Potentiodynamic Sensors

Modern dynamic electrochemical techniques offer additional enhancement of the information acquisition process, including selectivity and detection limit. Instead of holding the potential of the working electrode at a constant value, the potential is varied in some specific way. In that approach, we have a choice of several non-steady-state electrochemical techniques. They are all derived from the basic current–voltage concentration relationship (Section 5.1). A complete discussion of these electroanalytical techniques can be found in electrochemistry textbooks (Bard and Faulkner, 2001).

The most popular electroanalytical technique used at solid electrodes is Cyclic Voltammetry (CV). In this technique, the applied potential is linearly cycled between two potentials, one below the standard potential of the species of interest and one above it (Fig. 7.12). In one half of the cycle the oxidized form of the species is reduced; in the other half, it is reoxidized to its original form. The resulting current–voltage relationship (cyclic voltammogram) has a characteristic shape that depends on the kinetics of the electrochemical process, on the coupled chemical reactions, and on diffusion. The one shown in Fig. 7.12 corresponds to the reversible reduction of a soluble redox couple taking place at an electrode modified with a thick porous layer (Hurrell and Abruna, 1988). The peak current i_p is directly proportional to the concentration of the electroactive species C^* (mM), to the volume V (μL) of the accumulation layer, and to the sweep rate v (mV s^{-1}).

$$i_p = \frac{\beta n^2 F^2 V C^* v}{\Re T} \tag{7.26}$$

The constant β contains a partitioning coefficient of the analyte between the solution and the modifying layer, as well as the constants related to the bulk electrolysis in a small volume (i.e., thin) cell (Bard and Faulkner, 2001). If the electroactive species are confined to the electrode, if the couple is perfectly reversible, and if the extraction is fast on the time scale of the experiment, the peaks in the cyclic voltammogram occur at the same potential and the areas (charge) below the cathodic and anodic branches are equal, as is the case in Fig. 7.12. Obviously, any deviations from these conditions are reflected in the shape of the CV curve. Nevertheless, even then the relationship between the peak current i_p and the bulk concentration of the electroactive species can be reproducible. In the determination of Fe^{2+} using the above procedure, the linear calibration between 5×10^{-8} and 5×10^{-6} M concentration has been obtained.

If more electroactive species are present, the corresponding number of peaks is obtained. Moreover, when the standard potentials for the oxidation (and reduction) of those species are sufficiently separated ($\Delta E^0 > 100$ mV), the scanning of potential provides additional selectivity. Nevertheless, the same caveat regarding the relationship between selectivity and applied potential, as discussed above, applies. The overall selectivity of this sensor is due to the combination of two factors: selective complexation of Fe^{2+} with $2,2'$-bipyridile and the oxidation of the complex at the characteristic potential ($+1.0$ V vs. SCE). The value and the popularity of cyclic voltammetry and other electrodynamic techniques lie in the fact that they contain information not only about the concentration, but also about the chemical reactions accompanying the charge transfer.

7.5.3 Amperometric Biosensors

An effective means of achieving chemical selectivity is to modify the electrode with either a synthetic catalyst or an enzyme. Let us first consider the ampero-
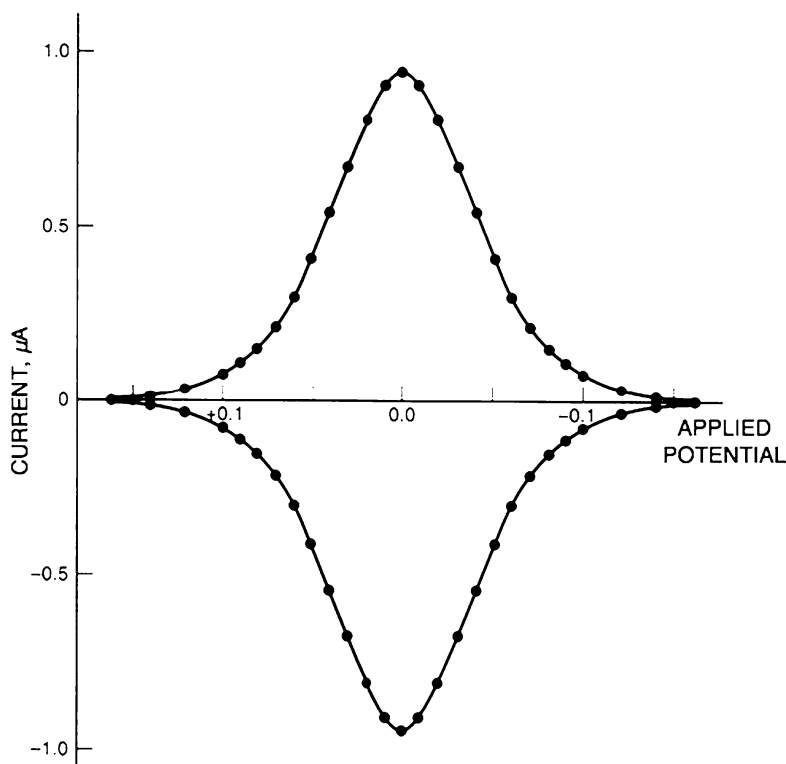
Fig. 7.12 Cyclic voltammogram obtained in a finite small volume ($V = 1.0\,\mu L$) at the scan rate $v = 1\,mV\,s^{-1}$ and concentration of the reversible analyte $C = 1.0\,mM$ (adapted from Bard and Faulkner, 2001, Section 11.7.3)

metric sensors that use enzyme layers. They are among the most representative and successful biosensors. The primary function of the enzyme is to catalyze the conversion of the electrochemically inactive substrate to some electroactive species, which can be monitored amperometrically. Thus, the general transduction is based on the selective conversion of the analyte to electroactive product **P** which is generated by the enzymatic reaction. Therefore,

$$S + E \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} ES \overset{k_2}{\rightarrow} P$$

$$P \pm ne \overset{E_1}{\Leftrightarrow} Q_1$$
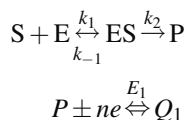
In principle, there is no requirement on how fast the electrochemical reaction should be because, as we have seen, it is usually possible to apply the working potential E_1 at which the kinetics of the charge transfer are fast. The origin of selectivity is in the enzymatic reaction.

In Fig. 2.10, the boundary between the enzyme-containing layer and the transducer has been considered as having either a zero or a finite flux of chemical species. In this respect, amperometric enzyme sensors, which have a finite flux boundary, stand apart from other types of chemical enzymatic sensors. Although the enzyme kinetics are described by the same Michaelis–Menten scheme and by the same set of partial differential equations, the boundary and the initial conditions are different if one or more of the participating species can cross the enzyme layer/transducer boundary. Otherwise, the general diffusion–reaction equations apply to every species in the same manner as discussed in Section 2.3.1. Many amperometric enzyme sensors in the past have been built by adding an enzyme layer to a macroelectrode. However, the microelectrode geometry is preferable because such biosensors reach steady-state operation.

There are many possibilities that can be implemented in enzymatic amperometric biosensors. Enzymatic schemes including several enzymes in one layer or multiple enzymatic layers in series can be used. If the cosubstrates are involved, it may be possible to oxidize or reduce one of them as well.

The enzyme can be incorporated into an amperometric sensor in a thick gel layer, in which case the depletion region due to the electrochemical reaction is usually confined within this layer. Alternatively, enzyme can be immobilized at the surface of the electrode or even within the electrode material itself, in which case the depletion region extends into the solution in the same way as it would for an unmodified electrode. In the latter case, the enzyme can then be seen as a true "electrocatalyst" that facilitates the interfacial electron transfer, which would otherwise be too slow. The general principles of the design and operation of these biosensors is illustrated on the example of the most studied enzymatic sensor, the glucose electrode (Fig. 2.14, Section 2.3.1).

There are several species in this reaction that can be used for electrochemical sensing. Detection of proton released from the gluconic acid was used in the potentiometric glucose electrode (Section 6.2.1). The amperometric sensor can be based on oxidation of hydrogen peroxide, on reduction of oxygen, or on the oxidation of the reduced form of glucose oxidase itself.

In addition to its independent importance, the oxygen Clark electrode has found major applications for biosensing, using enzyme oxidases as the biocatalysts. Undoubtedly, the top spot belongs to the glucose sensor. As of the writing of this text, an Internet entry "glucose sensor" netted over 33,000 citations of which more than 5,000 were electrochemical glucose sensors! The Clark electrode-based glucose sensor is a good example of a cosubstrate sensor, one substrate being oxygen and the other one glucose.

7.5.3.1 Oxygen-Based Enzyme Electrodes

The earliest design of the glucose electrode (which is generally applicable to any oxidase and uses oxygen as the ultimate electron acceptor) is based on the differential measurement of oxygen deficiency at the oxygen electrode, caused by the
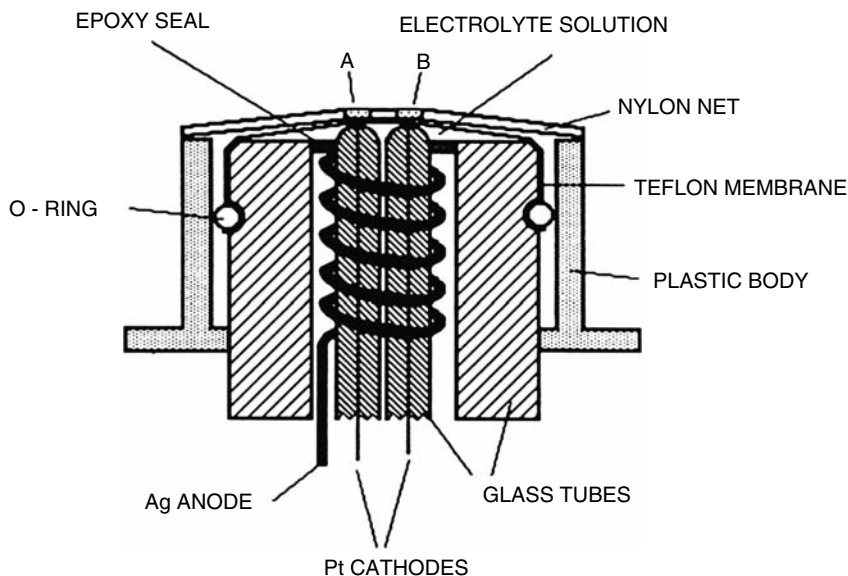
Fig. 7.13 "Dual" glucose sensor based on two oxygen electrodes, one (**a**) covered with active glucose oxidase and other (**b**) with deactivated glucose oxidase

enzymatic reaction. The principle of operation of this sensor is shown in Fig. 7.13. It consists of two identical oxygen electrodes: one (A) covered with the "active" glucose oxidase layer and the other one (B), containing the "deactivated" enzyme layer. The inactivation can be done either chemically, by ionizing radiation, or thermally. In the absence of the enzymatic reaction, the flux of oxygen to these electrodes, and therefore the diffusion limiting currents, are approximately equal $\Delta i = (i_{L1} \sim i_{L2})$. In practice, there is always some difference given by the unequal transport properties of the two layers, however, it is easy to account for this difference. The mathematical treatment that has been developed for this sensor is similar to that describing the hydrogen peroxide glucose electrode, which is discussed later. When glucose is present in solution, the enzymatic reaction takes place and the amount of oxygen reaching the surface of the "active" electrode is reduced by the amount consumed by the enzymatic reaction.

There are two major problems encountered with this approach. First is the self-inhibition caused by the hydrogen peroxide; this has been mentioned before. It is less acute in this case than in the hydrogen peroxide sensor because the catalase substantially eliminates any excess of H_2O_2. The second problem is encountered when the electrode is used in vivo. The tissue or blood concentration of oxygen, which is the cosubstrate, is low and at high glucose concentrations the current becomes limited by the availability of oxygen.

An elegant solution to this problem has been proposed (Gough et al., 1985). By design, the mass transport of oxygen has been increased, relative to that of glucose, by cylindrical diffusion into the enzyme layer and the transport of glucose restricted to the linear diffusion through the distal end of the sensor (Fig. 7.14). In other words,

the area through which oxygen reaches the electrode has been made larger relative to the area for the access of glucose. The most important outcome of this design has been the elimination of the dependency on oxygen. The sensor responds only to glucose even in the case of low concentration of oxygen in the bulk.

7.5.3.2 Oxidation of Hydrogen Peroxide

For this type of sensor, a catalase-free glucose oxidase must be used. In such a case, the hydrogen peroxide produced by the reaction with oxygen remains in the selective layer and can be detected by oxidation according to the reaction

$$H_2O_2 = O_2 + 2e^- + 2H^+$$
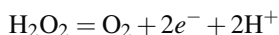
This sensor uses cylindrical microelectrode geometry (Fig. 7.14) for which the diffusion–reaction reaction is written in spherical coordinates, similar to (2.24).

$$\frac{\partial C_S}{\partial t} = D_S \frac{\partial^2 C_S}{\partial r^2} \pm \frac{\Re_{pH} v_{max} C_S}{(C_S + K_m)} \tag{7.27}$$

The sign in front of the reaction term is positive only for hydrogen peroxide. Also, the function $\Re_{pH}$ can be made constant by operating the sensor in a medium of high buffer capacity. This is clearly a distinct advantage compared to the potentiometric sensors in which the buffer capacity represented a major interference.

The initial and boundary conditions depend again on the model and on the operating conditions. Initially (at $t = 0$) there is no glucose (**G**) or hydrogen peroxide
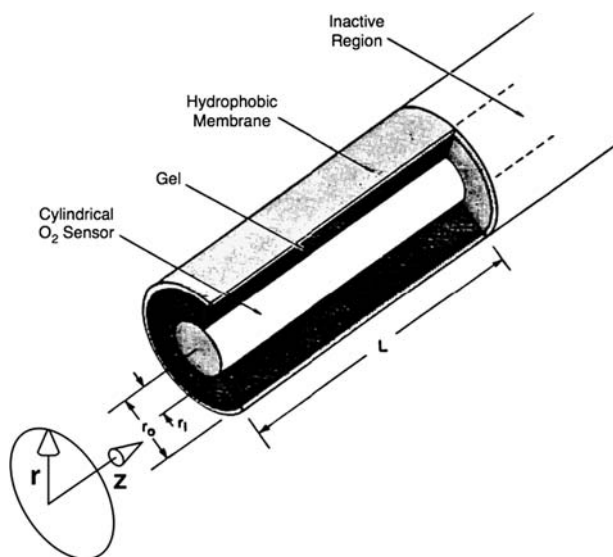
Fig. 7.14 Two-dimensional, cylindrical glucose sensor. Glucose diffuses to the sensor only in direction z whereas oxygen can reach the electrode by diffusion through the walls of the cylinder of length L

(**H**) present in the gel, which is saturated only with equilibrium concentration of oxygen (**O**).

$$C_G(r,0) = 0 \ \ for \ 0 < r < L$$
$$C_H(r,0) = 0 \tag{7.28}$$
$$C_O(r,0) = \alpha_O C_O^*$$

where $\alpha_O g$ is the partitioning coefficient for oxygen between the solution and the gel. After the reaction starts ($t > 0$), the boundary conditions are given by the following. There is no glucose present at the surface ($r = 0$) of the electrode and the reaction takes place entirely within the gel layer of thickness L. Thus, the boundary conditions for glucose are

$$C_G(0,t) = 0 \tag{7.29a}$$

and

$$\left(\frac{\partial C_G}{\partial r} \right)_{L,t} = 0 \tag{7.29b}$$

Oxygen, which is consumed by the primary glucose oxidation, is regenerated by the oxidation of hydrogen peroxide at the surface of the electrode. Moreover, because the operating potential is positive, oxygen is not consumed at the electrode. Therefore, the reaction is not limited by the concentration of oxygen for which the boundary conditions are

$$C_O(0,t) = C_O^* \tag{7.30a}$$

and

$$\left(\frac{\partial C_O}{\partial r} \right)_{L,t} = -\frac{D_H}{D_O} \left(\frac{\partial C_H}{\partial r} \right)_{L,t} \tag{7.30b}$$

The potential of the electrode is set in the diffusion limiting plateau region of hydrogen peroxide oxidation. Therefore the concentration of hydrogen peroxide at the surface of the electrode is also zero. Furthermore, there is no hydrogen peroxide in the bulk of the solution that is stirred. Therefore, its concentration at the gel/solution boundary is also zero. Thus, the boundary conditions for hydrogen peroxide are

$$C_H(0,t) = 0 \tag{7.30c}$$

$$C_H(L,t) = 0 \tag{7.31}$$

This means that H_2O_2 is generated only within the gel layer. Equation (7.27) can be solved exactly (Schulmeister and Scheller, 1985) for the above initial (7.28) and the boundary (7.29) conditions, yielding the explicit solution for the concentration profiles within the enzyme layer. The current is obtained by evaluating the concentration gradient of the hydrogen peroxide at the surface of the electrode, substituting it into Fick's law and multiplying it by the charge.

$$i_L = -2FAD_H \left(\frac{\partial C_H}{\partial r} \right)_{0,t} \tag{7.32}$$

It has been shown experimentally that the current is independent of the stirring rate, which means that the depletion layer is indeed confined to the interior of the gel layer.

A similar approach has been used to solve more complicated cases of two enzymes in one layer (Schulmeister and Scheller, 1985, p. 111) and the multi-layer/multienzyme model (Schulmeister, 1987, p. 223). It is important to note that oxidases are one of the largest group of enzymes and therefore the improved sensors for substrates other than glucose can and have been developed according to this scheme.

7.5.3.3 Direct Oxidation of Glucose Oxidase

Oxygen is the natural electron acceptor from the reduced form of GO, and for any other oxidase. The reduction of oxygen is not without problems. It leads to the degradation of the enzyme by the hydrogen peroxide and the dependency on available oxygen may create interferences in some applications, as has been discussed above. For these reasons, other electron acceptors have been investigated. The idea is to find some way by which the $(GO)_{red}$ could be reoxidized anaerobically, with the ultimate sink for the electrons being the electrode itself (Heller, 1999).

One possibility is to use low molecular weight mediators, which can shuttle electrons between the $(GO)_{red}$ and the surface of the electrode, thus bypassing the reaction with oxygen (see Fig. 7.9, Scheme 2). The obvious prerequisite for this scheme is the mobility of the mediator M_{Ox}, which must penetrate into the interior of the enzyme extract the electrons from the two flavin redox centers located deep inside the GO molecule (Fig. 7.15) and transport them to the electrode.

Thus, the following reactions take place in the enzyme layer,

$$FADH_2 + 2M_{ox} = FAD + 2H^+ + 2M_{Red}$$

and at the electrode,

$$M_{Red} = M_{ox} + 2e^-$$

Various fast redox couples such as ferrocene, ferro/ferri cyanide, and ruthenium hexamine have been used as mobile mediators. In order to be electron acceptors their standard potentials must be more positive than that of $FADH_2/FAD$ redox couple ($E^0 = 0.05\,V$, at pH $= 7$). The requirement of mobility is, however, in conflict with the lifetime of the sensor. Because the mediator is of comparable size to the substrate, it cannot be confined to the electrode proper by, for example, a dialysis membrane. In fact, the only way this type of sensor can operate is in a sample containing a sufficient concentration of the mediator (Cass et al., 1984). Obviously, this requirement makes such sensors suitable only for in vitro applications.
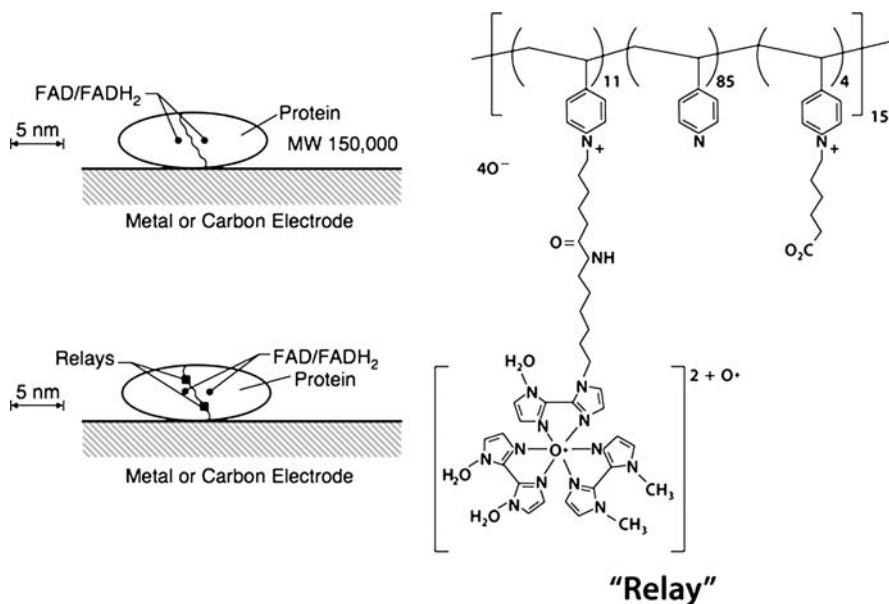
Fig. 7.15 Oxidase enzyme "wired" with electron transfer relays (adapted from Heller, 1999)

In principle, glucose oxidase could be oxidized directly at the electrode, which would be the ultimate electron acceptor. However, direct electron transfer between redox enzymes and electrodes is not possible because the $FADH_2/FAD$ redox centers are buried inside insulating protein chains (Heller, 1990). If it were not the case, various membrane redox enzymes with different standard potentials would equalize their potentials on contact, thus effectively shorting out the biological redox chains. The electron transfer rate is strongly dependent on the distance x between the electron donor and the electron acceptor.

$$k = \beta_1 \exp(-\beta_2 x) \tag{7.33}$$

The factor β_1 is a frequency factor and β_2 is a tunneling factor. Typical values of β_1 are 10^{12}–10^{13} s^{-1}, and $0.4 < \beta_2 < 0.13\,\text{nm}^{-1}$. Thus, the electron transfer rate drops by four orders of magnitude on increase of the distance from 8 to 17Å. The radius of glucose oxidase is 4.3 nm which means that communication between the electrode and the flavin centers inside the enzyme is not possible. Indeed, it has been found that GO is not electrochemically active on Pt, Au, carbon, and similar electrodes. The problem has been elegantly solved by Heller using an "enzyme wiring" concept. This approach involves forming redox hydrogel of cross-linked polyamine with pendant redox sites $(Os(bpy)_2Cl^{+/2+})$ and glucose oxidase (Fig. 7.15). Electrons are transferred from the flavin centers through the polyamine network acting as a molecular "wire" to the surface of the electrode. The cyclic voltammogram of this "wired" enzyme is shown in Fig. 7.16. The interference from molecules that
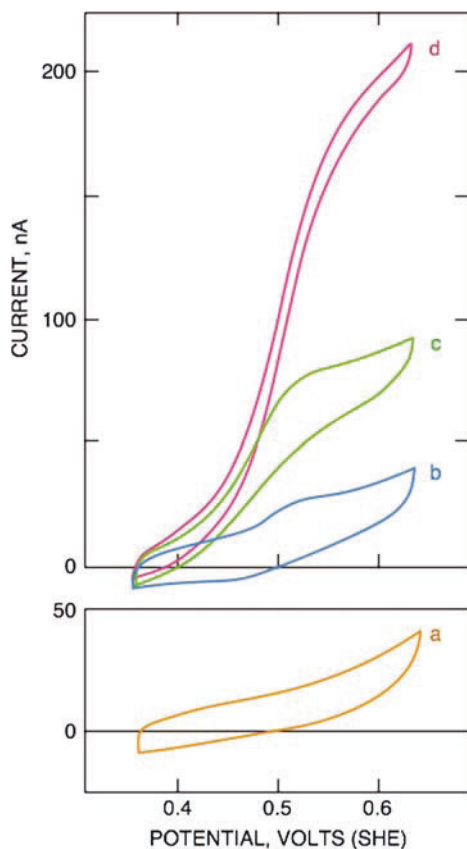
Fig. 7.16 Cyclic voltammograms of electrodes modified with (**a**) native GOD (or without GOD); (**b**) GOD containing 12 electron relays; (**c**) as in (**b**) but in the presence of 0.8 mM glucose; and (**d**) 5 mM glucose (adapted from Heller, 1999)

could react with the reactive intermediate, hydrogen peroxide, has been eliminated by a second layer containing catalase (Heller, 1999).

Other molecules have been used as the electron relays, such as ferrocene carboxylic acid and ruthenium pentamine $(Ru(NH_3)_5Cl_2)$. On average, one relay per 12,000–75,000 Da of enzyme is used. This approach seems to be also applicable to other oxidases. Therefore the amperometric response to glucose can be obtained even in the absence of oxygen or any other demonstrably soluble mediator at these electrodes. It is important to notice that electroactive glucose oxidase attached to the electrode fulfills the role of an electrocatalyst which selectively increases the rate of the electro-oxidation of glucose in preference to other substrates. In terms of the serial resistance model for the amperometric sensors discussed above (Fig. 7.8), it is the charge-transfer resistance for glucose that is selectively decreased. However, the signal itself comes from the mass transport resistance given by the geometry of the sensors. Thus, whichever approach proves to be ultimately more suitable for

Table 7.3 Examples of amperometric enzyme sensors

Substrate	Enzyme	Sensor	Stability	Response time	Range (M)
Glucose	Glucose oxidase	$Pt(H_2O_2)$	>14 months	1 min	$10^{-2}-10^{-4}$
		$Pt(O_2)$	>4 months	1 min	$10^{-1}-10^{-5}$
		$Pt(I^-)$	>1 months	2–8 min	$10^{-3}-10^{-4}$
L-Amino acids	L-AA oxidase	$Pt(H_2O_2)$	4–6 months	2 min	$10^{-2}-10^{-4}$
		$Gas(O_2)$		2 min	$10^{-2}-10^{-4}$
		$Pt(I^-)$	>1 months	1–3 min	$10^{-2}-10^{-4}$
Succinic acid	Succinic dehydrogenase	$Pt(O_2)$	<1 week	1 min	$10^{-2}-10^{-4}$
Aliphatic acid	Alcohol oxidase	$Pt(O_2)$	>4 months	30 s	$10^{-2}-10^{-4}$
Alcohols	Alcohol oxidase	$Pt(O_2)$	>4 months	30 s	$10^{-2}-10^{-4}$
Uric acid	Uricase	$Pt(O_2)$	4 months	30 s	$10^{-2}-10^{-4}$
Phosphate	Phosphatase/ glucose oxidase	$Pt(O_2)$	4 months	30 s	$10^{-2}-10^{-4}$

practical application, it is likely to be embodied in a microelectrode configuration, mainly for the reasons of its steady-state behavior.

The limited lifetime of the purified enzyme may be unacceptably short for a practical sensor, and such "purified" enzymes are too expensive or not even available. Mainly for these reasons, the amperometric enzyme electrodes employing cell cultures, tissues, and even whole organs have been prepared. Fundamentally, they are not different from the enzyme electrodes using purified enzymes immobilized in a synthetic matrix. In general, lifetime is indeed improved at the expense of a longer response time. The tissue or cell culture represents a diffusional barrier which is rather difficult to control. Sometimes the resistance of this barrier is such that the response of the sensor is lost. Thus, it has been found empirically that the maximum loading of bacterial cells should not exceed 1.5% of the total volume of the enzyme-containing layer (Kingdon, 1985; Marrese et al., 1987).

The three types of glucose electrode discussed here illustrate the major facets of design and operation of enzymatic amperometric sensors. Examples of amperometric enzyme electrodes for other substrates are shown in Table 7.3. The actual design details of these sensors depend on the enzyme kinetics involved and on the operating conditions under which they are used.

7.6 Amperometric Fuel Cells

If the number and type of gaseous species to be detected can be narrowly defined and if they remain invariable, it is possible to use a brute-force approach and to construct a chemical sensor without detailed knowledge of the mechanism of its operation. In such a case, the output of the sensor – be it voltage, current, or any other physical

parameter – is empirically optimized, to yield a calibration curve within the expected range of concentration of the species of interest. The main prerequisite is that this calibration is stable. Needless to say, the selectivity of these "fuel-cell-type" sensors is sometimes low. The relationship between the cell current and the concentration of chemical species of interest that is present in a more or less constant matrix can be empirically established and optimized. Therefore, their design has to be tailored to the specific application. Because the sample (usually an oxidizable species) enters into the cell through porous electrodes, similar to those used in fuel cells, the device is called a fuel cell sensor (although its purpose is not energy production).

If an electrochemical cell is constructed that contains a solvent/supporting electrolyte/electrode combination with as broad an electrochemical window as possible, a voltage can be applied at the working electrode such that one species can be oxidized (or reduced) with some degree of selectivity. The electrochemical process of interest can be further enhanced by electrocatalysis (c.f. Section 7.5). However, the comments that were made above, about the utility of applied potential as means of achieving selectivity, apply here. Therefore, the parameters that can be controlled and used to optimize the sensor performance in addition to applied potential are: solvent, electrolyte, additives, and electrode material. Additional and sometimes quite significant improvement in the performance is obtained by evaluating the output of the array of these sensors (see Chapter 10). Again, the most important requirement is the stability of their response. A major advantage of these sensors is the simplicity of their construction and the economy of their operation. Because the processes that affect the response are not fully known or controlled, it is often difficult to correct problems when they arise or to make rational design improvements (Fig. 7.17).

An example is the sensor for methane (Stetter and Li, 2008), which also shows sensitivity to several other oxidizable species. The cell current is calibrated against the concentration of methane in air of 75% relative humidity. The key element in this sensor is the low volatility electrolyte, such as γ-butyrolactone, propylene carbonate,
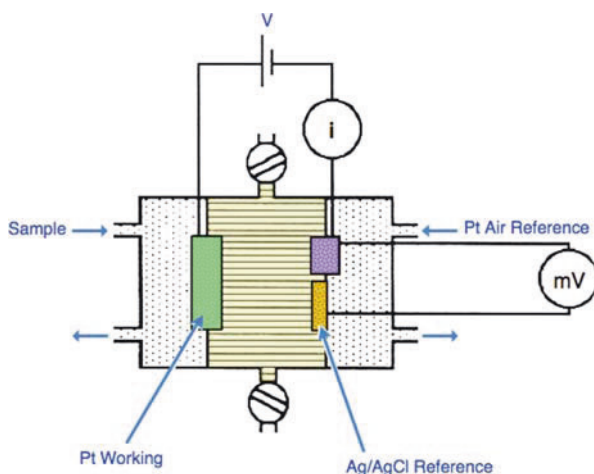
Fig. 7.17 Fuel cell amperometric sensor (adapted from Stetter and Li, 2008)

or sulfolane. All these solvents have a high boiling point (above 200°C) and a high anodic limit at the Pt electrode. Sodium or lithium perchlorate is usually used as the supporting electrolyte.

The porous platinum/Teflon electrodes separate the electrolytic cell from the gaseous reference chamber on one side and the sample chamber on the working electrode side. The applied voltage is controlled by a potentiostat. The sample enters into the electrolytic cell through the porous electrodes, the pore size of which also needs to be closely controlled in order to prevent their flooding with the solvent. An example of an electrochemical reaction of interest is oxidation of methane under conditions of humid air.

$$CH_4 + 2H_2O \Rightarrow CO_2 + 8H^+ + 8e^-$$
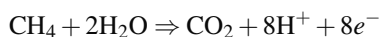
Using an 800 mV applied voltage, the response of the sensor has been evaluated in the range of 12–100% CH_4 in air (Fig. 7.18). Major interferences have been caused by the presence of nitrous oxide, ethane, hydrogen, and carbon monoxide. Other sensors of this type have been described, but they differ only in the details of the design and not significantly in the concept.

Introduction of room-temperature ionic liquids (RTIL) as electrochemical media promises to enhance the utility of fuel-cell-type sensors (Buzzeo et al., 2004). These highly versatile solvents have nearly ideal properties for the realization of fuel-cell-type amperometric sensors. Their electrochemical window extends up to 5 V and they have near-zero vapor pressure. There are typically two cations used in RTIL: N-dialkyl immidazolium and N-alkyl pyridinium cations. Their properties are controlled mostly by the anion (Table 7.4). The lower diffusion coefficient and lower solubility for some species is offset by the possibility of operation at higher temperatures.

7.7 High Temperature Limiting Current Sensors

Yttrium stabilized zirconia (ZrO_2–Y_2O_3) as an electrolyte for reduction of molecular oxygen at elevated temperatures (400–800°C) has been already discussed in Section 6.2.3.4. In fact, both the reduction of oxygen and the oxidation of the oxide ion at the Pt/zirconia interface is reversible and the transport of both species in zirconia is so rapid it is possible to construct an electrochemical oxygen pump, which is the heart of the limiting current oxygen sensor described in this chapter (Saji, 1987). The overall electrochemical reaction that takes place at the porous Pt electrode is

$$O_2 + 4e^- + 2V_O^+ \Leftrightarrow 2O^-$$
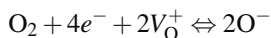
where V_O^+ is the oxygen vacancy in the ZrO_2 lattice. This layer is sandwiched between two porous (1 μm thick) Pt or Pd electrodes at the bottom of an hermetically sealed cavity (Fig. 7.19). As oxygen is reduced at the cathode, oxide ions diffuse through the 500 μm thick zirconia to the anode where they are oxidized to
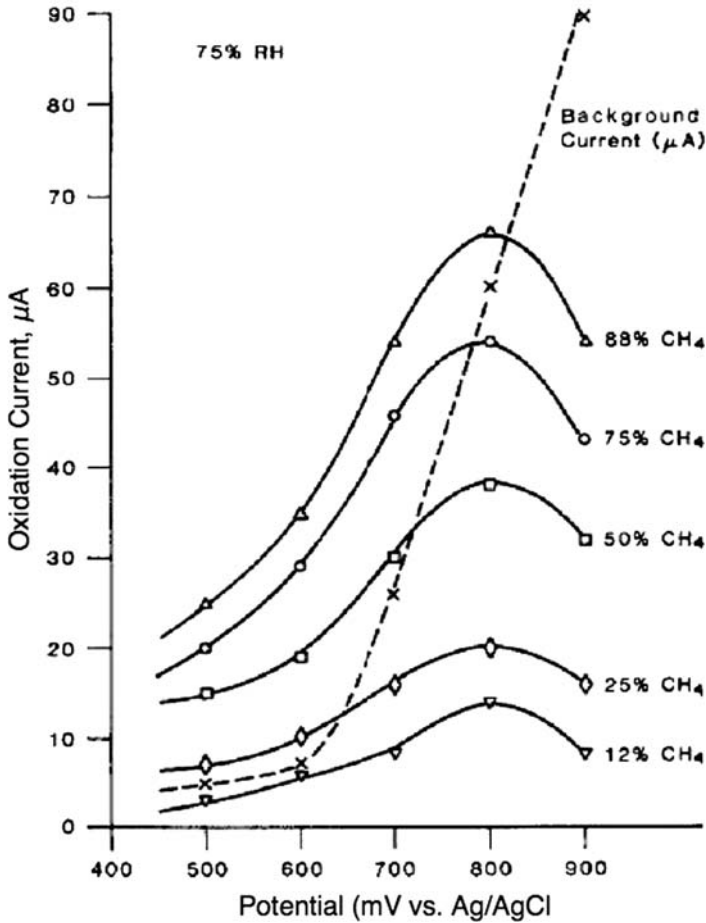
Fig. 7.18 Response of fuel cell amperometric sensor to methane in air (adapted from Stetter and Li, 2008)

oxygen, which then exits to the atmosphere through the porous anode. Thus, molecular oxygen is pumped from inside the cavity, through ZrO_2, to the outside. The electrical current that flows through the circuit is limited by the oxygen transport to the surface of the cathode. In order to reach the mass transport limit, it is necessary to place a diffusional barrier in series with the oxygen pump. In practical sensors, this is done in two ways: either a small diameter hole is drilled in the cover (Fig. 7.19a) or the pump is covered with a porous layer of, for example, $MgO–Al_2O_3$ (Fig. 7.19b). Thus, the term "limiting current sensor" refers here to the diffusion barrier in series with the oxygen pump and not to the diffusion limiting process in the normal electrochemical sense. Nevertheless, the principles governing the operation of this sensor are the same as those applicable to liquid-phase sensors.

Table 7.4 Values of diffusion coefficients and solubilities for (A) oxygen and (B) carbon dioxide in various RTIL compared with conventional solvents. $Py_{14} = N$-methyl butyl pyrrolidinium; $P_{14,666} = $ tris(n-hexyl)tetradecylphosphonium; Tf = triflate; MIM = methyl imidazolium; FAP = fluoroethyl phosphate (adapted from Buzzeo et al., 2004)

Solvent	Diffusion coefficient $(\times 10^{-1} \text{m}^2 \text{s}^{-1})$	Solubility
A. Oxygen		
H_2O	16	1.3
DMSO	21	2.1
DMF	47.5	4.8
McCN	110	8.1
$[C_2MIM][N(Tf)_2]$	7.3	3.9
$[N_{6222}][N(Tf)_2]$	1.5	11.6
$[C_4MIM][BF_4]$	Not available	7.0
$[C_4MIM][PF_6]$	2.2	3.6
$[Py_{14}][N(Tf)_2]$	5.2	6.1[b]
$[P_{14,666}][N(Tf)_2]$	7.5	6.0[b]
$[P_{14,666}][FAP]$	6.1	7.8[b]
B. Carbon dioxide		
H_2O	24	33
DMSO	10	125
DMF	35.6	198
McCN	383	280
$[N_{6222}][N(Tf)_2]$	2.3	55
$[C_4MIM][BF_4]$	Not available	30
$[C_4MIM][PF_6]$	Not available	90.2

For the hole of the diameter larger than the mean free path of O_2 diffusion, the flux of oxygen, J_O, is given as

$$J_O = -D_O A \frac{dC_O}{dx} + J_c \frac{C_O}{C} \qquad (7.34)$$

where D_O is the molecular diffusion coefficient of oxygen in the gas phase, C is the total gas concentration, C_O is the bulk oxygen concentration, and A is the cross-section of the aperture. The transport is considered to be unidirectional, with $x = 0$ being at the orifice of the hole. The length of the hole is L. Because the aperture is the only opening in the cavity, the electrochemical reduction causes pumping action and a gas flow, such that $J_c = J_0$. Equation (7.34) can be solved for the boundary conditions: at $x = 0$, $C_O = C_O(0)$; at $x = L$, $C_O = C_O(L)$. Considering these conditions and the proportionality between the oxygen flux at the electrode and the current (7.3), we obtain:

$$i = 4FD_O A \frac{C_O}{L} \qquad (7.35)$$

Equation (7.35) yields the expression for steady-state current:
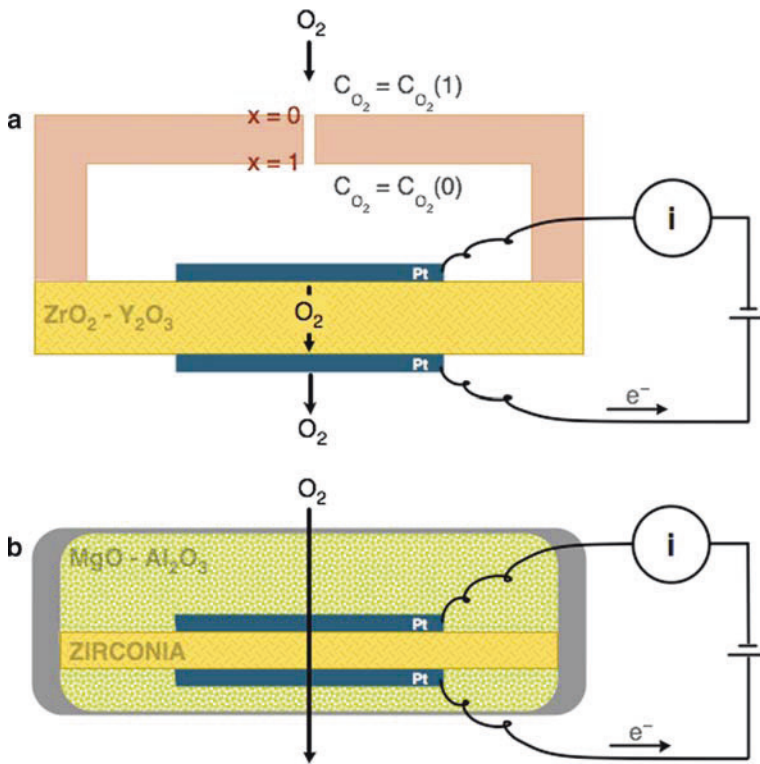
Fig. 7.19 High temperature limiting current sensor: (**a**) "pinhole" version and (**b**) porous "diffusional barrier" version

$$i = \frac{4FD_O AC}{L} \ln \frac{1 - C_O(L)/C}{1 - C_O(0)/C} \tag{7.36}$$

When the oxygen concentration inside the cavity reaches zero (i.e., $C_O(L) = 0$) the current becomes mass transport limited; that is, it reaches its limiting current value and (7.36) becomes

$$i_L = \frac{4FD_O AC}{L} \ln[1 - C_O(0)/C] \tag{7.37}$$

Using the approximation $\ln(1 - x) \sim -x$, for small values of $C_O(0)/C$, (7.36) becomes

$$i_L = \frac{4FD_O AC_O(L)}{L} \tag{7.38}$$

Thus, for small values of partial pressure of oxygen in the ambient gas, the limiting current is directly proportional to the oxygen concentration. This approximation

holds up to approximately 20% partial pressure of oxygen (Pham and Glass, 1997). At higher oxygen concentrations, (7.36) has to be used.

The complete current–voltage characteristics of the sensor can be derived from the similar consideration that was used for derivation of the *i–E* curve for liquid electrolytes. Because the potentials at each electrode are reversible, their difference can be expressed by the Nernst equation for the concentration of oxygen at the anode $C_O(0)$ and at the cathode $C_O(L)$. The current flowing through the layer generates a voltage drop iR_b, where R_b is the bulk resistance of the ZrO$_2$ layer.

$$V = iR_b + \frac{\Re T}{4F} \ln \frac{C_O(0)}{C_O(L)} \tag{7.39}$$

Substituting for *i* from (7.36) gives

$$V = \frac{4FD_OACR_b}{L} \ln \frac{1 - C_O(L)/C}{1 - C_O(0)/C} + \frac{\Re T}{4F} \ln \frac{C_O(0)}{C_O(L)} \tag{7.40}$$

This equation can be solved numerically. At concentrations of oxygen below 0.2%, the limiting current plateau becomes narrow and the onset of another electrode process contributes to positive error. It has been attributed to the reduction of CO$_2$ or to the increase of unspecified electronic current (Pham and Glass, 1997). Excellent agreement between the calculated and experimental values has been observed for a wide range of temperatures (Fig. 7.20).

Operation of the sensor with porous diffusional barrier (Fig. 7.19b) is similar except that the constants have slightly different physical meanings. In that case, A_{eff} is the area of the cathode and *d* is the thickness of the porous layer. For a low concentration of oxygen
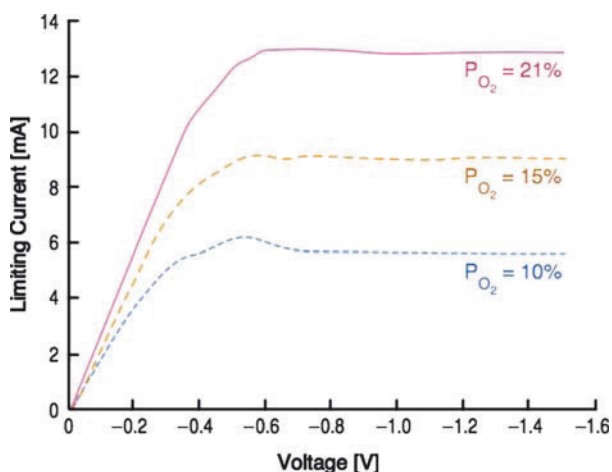
Fig. 7.20 Response of limiting current oxygen sensor at 700°C to various partial pressures of oxygen (adapted from Pham and Glass, 1997)

$$i_{\text{L}} = \frac{4FD_{\text{O,eff}}A_{\text{eff}}C_{\text{O}}(d)}{d} \tag{7.41}$$

Here, the length L in (7.38) has been replaced by porous layer thickness d and the surface area A_{eff}. The effective diffusion coefficient $D_{\text{O,eff}}$ characterizes the transport through porous medium and includes both regular diffusion and the Knudsen diffusion coefficient $D_{\text{O,Kn}}$, which has a different temperature dependence from diffusion in bulk.

$$1/D_{\text{O}},\text{eff} = 1/D_{\text{O}} + 1/D_{\text{O,K}} \tag{7.42}$$

The relative contribution of these two processes is controlled by the average pore size. It is therefore possible to optimize the temperature sensitivity by the judicious choice of the porosity of the diffusion barrier layer (Vacek et al., 1986). The temperature coefficient of the sensor with 0.1–0.2 μm average pore diameter is 0.04%/°C between 640°C and 800°C.

The background current of the sensor is given by the contribution to the current of the electronic conductivity through the zirconia layer. It can be expressed as the equivalent oxygen concentration causing the same current. At 700°C, it is approximately equivalent to 10^{-40} atm of oxygen. The contribution of the flow through the pinhole to the mass transport does not exist in the porous layer sensor and neither does the time constant associated with the cavity on the cathode side. For all these reasons, the porous layer version of the high temperature limiting current sensor has better performance characteristics than the pinhole version. It is also less susceptible to error due to inadvertent plugging of the pinhole with some particulate matter.

By far the most important practical use of this sensor is for automotive applications, namely for the control of the air to fuel ratio. It compares favorably with the surface conductivity or high temperature potentiometric sensor (Logothetis, 1987). Other gases could be detected on the same principle provided that the right materials for the electrochemical pump were used. The electrode materials/solid electrolytes used for the construction of potentiometric high temperature sensors (see Table 6.7) could serve as guidance.

Food for Thought #7

Electrode Placement in Potentiometric and Amperometric Measurement

For in vivo electrochemical measurement, the placement of the electrodes is important. In the potentiometric measurement (e.g., measurement of the pH of gastric juices) both the working electrode and the reference electrode are placed in the stomach. On the other hand, when amperometric measurement is done, the working electrode is in the place where you are getting the information from, but the auxiliary (reference) electrode can be anywhere, even on the skin.

1. Explain why this is so.
2. What will happen when you place the reference electrode on the skin in a potentiometric measurement?
3. What will happen when the electrodes are reversed in amperometric measurement?

Constant Current

4. The transduction principle in amperometric sensors is proportionality between the bulk concentration of the analyte and the mass transport limited current. There are two ways to keep the proportionality "constant" really constant. What are they?

Fuel Cell Amperometric Sensors

5. The working and auxiliary electrodes in this sensor are placed in the same compartment. In what type of reactions could such an arrangement cause real problems?

Symbols

A	Area of the electrode
C_O^*	Bulk concentration
$C_O(0,t)$	Surface concentration
d	Diameter
D	Diffusion coefficient
$D_m S_m$	Oxygen transmissivity of membrane m
E	Potential
erf	Error function
F	Faraday constant (96,493 coulombs)
f	Activity coefficient
I (or i)	Current
J	Flux
j	Current density
k	Rate constant
M	Mobility
m_O	Mass transport coefficient of species O
n	Number of electrons
P	Partial pressure
R	Resistance
$\Re_{pH}$	Dimensionless pH-dependent activity of enzyme

r_o	Radius of the hemispherical electrode
S	Solubility
t	Time
U	Linear velocity
v	Sweep rate
V	Volume
x	Distance
α	Partitioning coefficient
$\varepsilon_\%$	Accuracy
π	Mathematical constant *pi* (3.1415...)
δ	Thickness
ν	Kinematic viscosity

References

Bard, A.J. and Faulkner, L.R. (2001) *Electrochemical Methods, Fundamentals and Applications*. Wiley.

Buzzeo, M.C., Hardacre, C., and Compton, R.G. (2004) *Anal. Chem*. 76, 4583–4588.

Cass, A.E.G., Davis, G., Francis, G.D., Hill, H.A.O., Aston, W.J., Higgins, J.I., Plotkin, E.V., Scott, L.D.L., and Turner, A.P.F. (1984) *Anal. Chem*. 56, 667.

Fatt, I. (1976) *Polarographic Oxygen Sensor*. CRC Press.

Fleischmann, M., Pons, S., Rolison D.R., and Schmidt, P.P. (1987) *Ultramicroelectrodes*. Datatech Systems Inc.

Gough, D.A., Lucisano, J.Y., and Tse, P.H.S. (1985) *Anal Chem*. 57, 2351.

Heller, A. (1990) *Acc. Chem. Res*. 23, 128.

Heller, A. (1999) *Annual Review of Biomedical Engineering*, vol. 1. pp. 153–175.

Hitchman, M.L. (1978) *Measurement of Dissolved Oxygen*. Wiley.

Hurrell, H.C. and Abruna, H.D. (1988) *Anal. Chem*. 60, 254.

Kingdon, C.F.M. (1985) *Appl. Microbiol. Biotechnol*. 21, 176.

Koryta, J., Dvorak, J., and Kavan, L. (1993) *Principles of Electrochemistry*. Wiley.

Kreuzer, F., Kimmich, H.P., and Brezina, M. (1980) Polarographic determination of oxygen in biological materials. In: Koryta, J. (Ed.) *Medical and Biological Applications of Electrochemical Devices*. Wiley.

Logothetis, E.M. (1987) In: Turner, D.R. (Ed.) *Oxygen Sensors for Automotive Applications, in Chemical Sensors*. Electrochemical Society, p. 142.

Mano, N., Fernandez, J.L., Kim, Y., Shin, W. Bard, A.J., and Heller, A. (2003). *J. Am. Chem. Soc*. 125, 15290.

Marrese, C.A. Miyawaki, O., and Wingard, Jr., L.B. (1987) *Anal. Chem*. 59, 248.

Pham, A.Q. and Glass, R.S. (1997) *J. Electrochem. Soc*. 144, 3929.

Saito, Y. (1967) *J. Appl. Physiol*. 23, 979.

Saji, K. (1987) *J. Electrochem. Soc*. 134, 2430.

Schulmeister, T. (1987) *Anal. Chim. Acta*. 198, 223.

Schulmeister, T. and Scheller, F. (1985) *Anal. Chim. Acta*. 170, 279.

Soukharev, V., Mano, N., and Heller, A. (2004) *JACS* 126(27), 8368.

Stetter, J.R. and Jing, Li. (2008) *Chem. Rev*. 108, 352.

Vacek, V., Linek, V., and Sinkule, J. (1986) *J. Electrochem. Soc*. 133, 540.

Wightman, R.M. and Wipf, D.O. (1988) Ultramicroelectrodes. In: Bard, A.J. (Ed.) *Electroanalytical Chemistry*, vol. 15. Marcel Dekker.

Chapter 8
Conductometric Sensors

8.1 Introduction

In Chapters 6 and 7, we discussed potentiometric and amperometric sensors, respectively. The third basic electrochemical parameter that can yield sensory information is the conductance of the electrochemical cell (Fig. 8.1). Conductance is the reciprocal of resistance. It is related to current and potential through the generalized form of Ohm's law (C.1). If the measurement is done with AC signal conductance (G) becomes frequency-dependent conductance $G(\omega)$ and the resistance R becomes impedance $Z(\omega)$.

$$G(\omega) = \frac{I(\omega)}{E(\omega)} \tag{8.1}$$

The unit of conductance is Siemens (mho), which is the reciprocal of resistance R (ohm). Resistance is related to the resistivity (ρ) by the ratio of the length (L) and cross-section (A).

$$R = \rho \left(\frac{L}{A}\right) \tag{8.2}$$

On the other hand, conductivity (σ) or specific conductance (reciprocal of resistivity ρ) is a material property that is normalized with respect to area, potential gradient, and time. It is expressed as the ratio of current density j (A cm^{-2}) and electric field E (V cm^{-1}).

$$\sigma = \left(\frac{j}{E}\right) \tag{8.3}$$

Its unit is Siemens^{-1} cm^{-1}. These terms are used throughout this section interchangeably. These sensors are popularly called "chemiresistors," which clearly implies their function. They are simple to fabricate, but the interpretation of their responses and the mechanism of their operation are anything but simple. The fundamental equivalency between current and resistance, which is inherent in the generalized Ohm's law (8.1), sometimes blurs the line between amperometric and conductometric sensors. Nevertheless, it is important to remember that chemiresis-
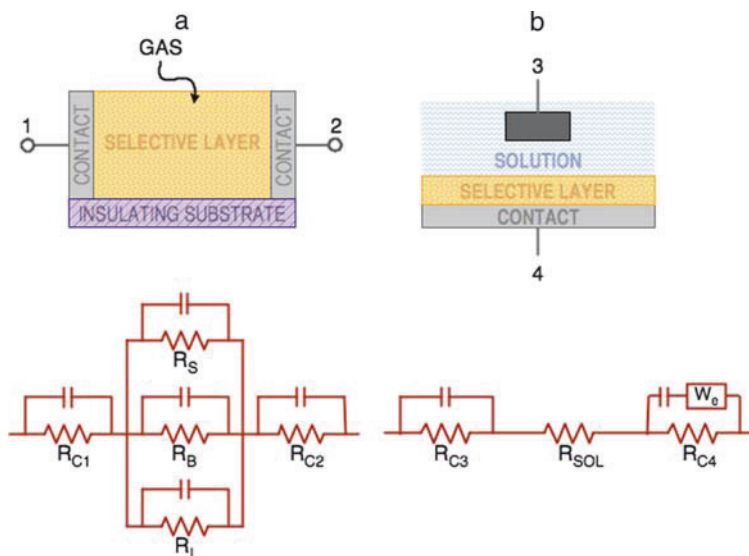
Fig. 8.1 General chemiresistors and their equivalent circuits: (**a**) lateral configuration in which any of the five resistances can be modulated by chemical interaction; (**b**) impedimetric chemiresistor in which capacitance C_B is chemically modulated

tors are fundamentally electrochemical devices, and that both the interpretation of the results and the optimization of their design should be done with the help of the general rules governing the operation of electrochemical cells.

8.1.1 The Origin of Sensor Response

Conductometric sensors are usually bipolar devices. However, they come in different configurations with different corresponding equivalent circuit diagrams, as shown in Fig. 8.1. The sample interacts with the selective layer, which is arranged in the lateral configuration (Fig. 8.1a). In this type of conductometric sensor, the measurement is done in DC mode. Because the conducting samples would short out the resistance of the chemiresistors, the domain of their application lies in the gas phase or in nonconducting liquids. In the second arrangement, which corresponds to a regular electrochemical cell (Fig. 8.1b), the impedance is measured perpendicular to the interface between the selective layer and the conducting ionic sample and the measurements are done mostly in an AC mode with a periodically changing excitation signal.

Discussion of general operation of conductometric sensors can begin by analyzing Fig. 8.1. Once again we realize that the electrochemical cell is a complex arrangement of resistances and capacitances. The primary interaction between the sample and the sensor involves the selective layer; the sensory information is

obtained by the selective modulation of only one of those equivalent circuit elements. Thus, we have to consider the contact resistance (R_c) between the electrodes and the selective layer, the bulk resistance (R_b) and the surface resistance (R_s) of the selective layer, and the interface resistance (R_i) which is located at the interface between the selective layer and the insulating substrate. The combination of these resistances yields an equivalent resistance that is measured between points 1 and 2 (Fig. 8.1a) or points 3 and 4 (Fig. 8.1b).

Because resistance is an intensive property (i.e., size-dependent), it is possible to localize the source of the signal at one electrode by making its contact area with the selective layer smaller than that of the other electrode. This is exactly the same argument as we have used in the discussion of the amperometric sensing circuit (Chapter 7).

Turning to Fig. 8.1a, we can dissect the overall impedance between points A and B into partial impedances (Z) and admittances (Y) and represent them with an equivalent impedance circuit. The overall impedance of this circuit is then

$$Z = \bar{Z}_C \mathrm{d}n_C + (\bar{Y}_S \mathrm{d}n_S + \bar{Y}_B \mathrm{d}n_B + \bar{Y}_I \mathrm{d}n_I)^{-1} \tag{8.4}$$

where the number of moles of chemical species interacting with each node affects its partial admittance ($\bar{Y}$) or impedance ($\bar{Z}$). (For a definition of partial properties, see Appendix A.)

$$\bar{Y}_i = \left(\frac{\partial Y}{\mathrm{d}n_i}\right)_{n_{j \neq i}} \text{ and } \bar{Z}_i = \left(\frac{\partial Z}{\mathrm{d}n_i}\right)_{n_{j \neq i}} \tag{8.5a}$$

and

$$Y_i = Z_i^{-1} = f(C_i, E_i) \tag{8.5b}$$

The same consideration applies to the impedance measurement according to Fig. 8.1b. It is a normal electrochemical interface to which the Warburg element (Z_W) has been added. This element corresponds to resistance due to translational motion (i.e., diffusion) of mobile oxidized and reduced species in the depletion layer due to the periodically changing excitation signal. This refinement of the charge-transfer resistance (see (5.23), Chapter 5) is linked to the electrochemical reaction which adds a characteristic line at $45°$ to the Nyquist plot at low frequencies (Fig. 8.2)

$$Z_W = R_{ct}\left(1 + \frac{2k_0}{\sqrt{j\omega D}}\right) \tag{8.6}$$

Equation (8.6) is a simplified approximation. For an exact definition, please see, for example, Bard and Faulkner (2001). It is mentioned here because it appears at low frequency (0–20 Hz) excitations. At higher (>100 Hz) frequencies, the second term can be neglected and the Nyquist plot reverts to the simple semicircle.

The discussion of the origin of the signal helps in the classification of conductometric sensors. Chemiresistors (Fig. 8.1a) are normally measured in DC mode and the capacitances in the equivalent circuit can be neglected. Thus, the analytical information is obtained from the modulation of surface (R_S), contact (R_C), or
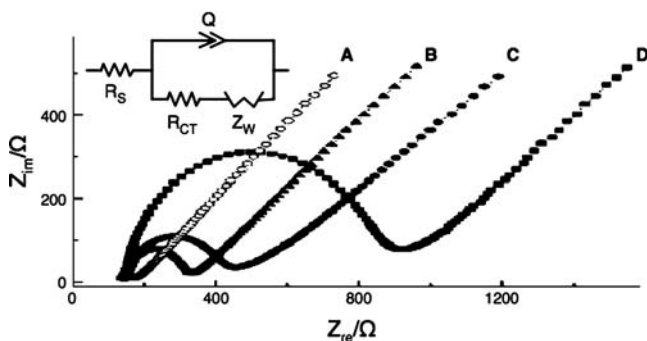
Fig. 8.2 Nyquist plot of "impedimetric sensor" (Section 8.4.3.3). The redox couple ferro/ferricyanide recorded at (**a**) bare Au electrode; (**b**) Au electrode modified with thrombin-binding aptamer; (**c**) Au/aptamer electrode with added 2-mercapto ethanol; and (**d**) electrode with bound complementary thrombin/aptamer complex. Each plot was obtained separately (adapted from Radi et al., 2005)

bulk (R_B) resistances. Chemiresistors utilizing organic semiconductors operate on the principle of chemical modulation of bulk and/or resistance. The interfacial resistance (R_I) can play a role in the overall response, but is normally not used as a source of the signal. Chemiresistors based on inorganic semiconductors usually depend on chemical modulation of surface resistance (R_S). In contrast, the chemical information in impedance sensors (Fig. 8.1b) is obtained from the modulation of the bulk resistance (R_B) or even capacitance (C) of the selective layer.

8.1.1.1 Contact Resistance

In most conductometric sensors reported to date, the electrodes are metallic which means that they are electronic conductors. This limits the number of types of possible contacts to: metal–metal (electronic conductor), metal–insulator–metal, metal–semiconductor, and metal–insulator–semiconductor. The metal–metal contact is generally ohmic and the contact resistance (R_C) is not subject to chemical modulation if polarization is avoided. Polarization can happen when either a small density of the DC current or a small AC excitation amplitude is used and the selective layer has a high concentration of charge carriers. Also, a high concentration of carriers results in high conductance (or low resistance R_B) of the selective layer. Hence, the change in conductivity due to interactions with the analytes leads only to a very small relative signal change, $\Delta R/R_B$. These prerequisites are based on the general electrochemical principles discussed in Chapter 5. Unfortunately, many metal/selective layer junctions are nonohmic due to the presence of a thin insulating layer. The result is a partially blocked contact. Another cause is the presence of a Schottky barrier (Appendix C). Here, the low-level doping of the semiconductor results in formation of a depletion layer, which results in a mismatch of the Fermi levels. These mechanisms of charge injection are discussed below.

8.1.1.1.1 Partially Blocked Contacts

The current density–voltage relationship of a contact in which a thin layer of insulating oxide covers the metal and the selective layer is electronically conducting follows the Fowler–Nordheim equation, which describes the tunneling current at metal–insulator–metal junctions.

$$I = B - (\phi_B + V) \exp\left\{ -4\pi d \left[\frac{h(\phi_B + V)}{2me} \right]^{1/2} \right\} \tag{8.7}$$

Here V is the applied voltage, ϕ_B is the average barrier height, d is the insulator thickness, h is Planck's constant, m is the mass, e is the charge of electron, and B is a constant dependent on barrier height and thickness. Both barrier height and the thickness of the oxide can be chemically modulated, giving rise to chemiresistive behavior. Such barrier junctions apparently exist in most chemiresistors, which do not use noble metals for the contact.

8.1.1.1.2 Schottky Contacts

If the selective layer is a semiconductor, a Schottky barrier may exist at the insulator-free contact, which also can give rise to the chemically modulated resistance. Chemically sensitive Schottky barrier contacts can be formed between the metal and the Organic Semiconductor (OS). The current through the Schottky diode is a highly nonlinear function of the applied potential (Fig. 8.3a) and can be used for chemical sensing (Potje-Kamloth, 2008). The measurement is performed by applying constant current through the junction and evaluating the shift of the bias voltage upon
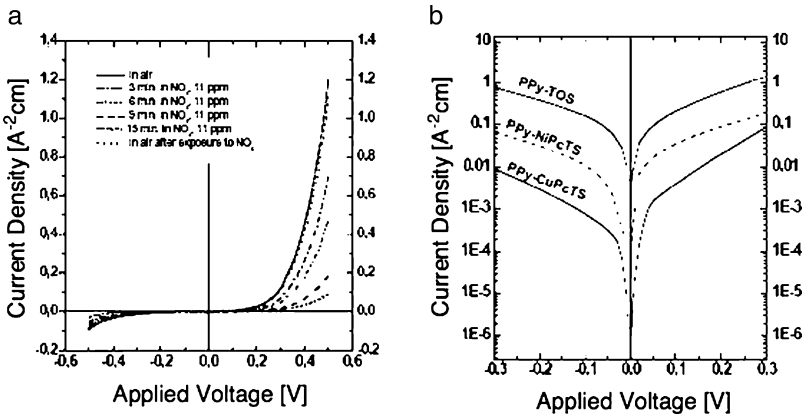
Fig. 8.3 (**a**) Schottky diode characteristic modulated by the electron acceptor gas (NO_2); (**b**) effect of different primary dopants on the diode characteristics (adapted from Nguyen and Potje-Kamloth, 1999)

exposure to the secondary dopant (Fig. 8.3b). Other insulator-free barrier contacts, called heterojunction contacts, can be formed between two dissimilar semiconductors, for example, OS and an inorganic semiconductor such as silicon. In either type, the OS can be doped with different anions, which affects the response time and sensitivity (Fig. 8.3b; Nguyen and Potje-Kamloth, 1999; Potje-Kamloth, 2001).

Contacts made of noble metals can be prepared essentially oxide-free. In that case, the junction current for $V >> kT/e$ is given by

$$J = A_R T^2 \exp \left[\frac{-(\phi_B - eV)}{kT} \right] \tag{8.8}$$

where A_R is the Richardson constant for thermionic emission of electrons, k is a rate constant, V is the bias voltage, T is the temperature, and e is the elementary charge.

$$A_R = 4\pi mek^2/h^3 \tag{8.9}$$

The saturation current (at $V = 0$) is

$$J_0 = A_R T^2 \exp \left[\frac{-\phi_B}{kT} \right] \tag{8.10}$$

Different complications arise if the selective layer is a solid-state ionic conductor. At such an interface, a net electrochemical reaction, governed by Faraday's law, takes place and the mass transport of the electroactive ionic species within the contact region and formation of a depletion layer must be considered. In general, when the response of the sensor depends on the chemical modulation of the contact resistance by one of the above mechanisms it will be a strongly nonlinear function of concentration. Furthermore, because R_c is always dependent on the applied voltage, the optimization of the response must be done by examining the voltage–current characteristics of the contact.

In principle, the degradation of the junction can be avoided if the excitation is done with a small amplitude (~10 mV) AC voltage. In that case, the electrochemical degradation due to the Faradayic reaction is reversed within one cycle of the applied voltage. For a time-varying excitation signal (e.g., sinusoidal voltage or current) the capacitances in the equivalent electrical circuit must also be taken into consideration. The contact capacitance (C_c), which is analogous to the double-layer capacitance of the electrolytic cell, may change as a result of the accumulation of the charged species at the contact. Also, the bulk capacitance (C_B) of the selective layer may change as a result of the change of the bulk dielectric constant, due to the uptake of some chemical species of different dipole moment, for example, water. The depletion-layer capacitance (C_s) is affected by the density of the surface states and by the dielectric constant of the depletion region and is dependent on the current. If the contact resistance is lower than the resistance of the selective layer, it can be neglected. The modulation of the bulk conductivity or of the surface conductivity of the selective layer is then the source of the signal.

8.1.1.2 Bulk Resistance

The bulk conductivity σ depends on the concentration of charge carriers and on their mobility, either of which can be modulated by exposure to the gas. The first prerequisite of such an interaction is the penetration of the analyte to the interior of the layer. The second is the ability of the gas to form a charge-transfer complex with the selective layer. This process then constitutes secondary doping, which affects the overall conductivity. For a mixed semiconductor, the overall conductivity is determined by the combined contribution from the holes (p) and electrons (n), as given by the general conductivity equation.

$$\sigma = e\mu_p C_p + e\mu_n C_n \tag{8.11}$$

Of course, in case of the intrinsic semiconductor for which $C_n = C_p = C_i$ (8.11) simplifies to

$$\sigma = eC_i(\mu_p + \mu_n) \tag{8.12}$$

where μ_p and μ_n are the mobilities of the positive and negative charge carriers, respectively. Equation (8.12) applies to conductivity in a homogeneous medium in which the translational mobility of the charge carrier is determined by molecular friction and through the Stokes–Einstein equation (B.12) can be related to diffusion coefficient D. However, if the charge moves by hopping from one redox site to another, the Dahms–Ruff equation applies (Bard and Faulkner, 2001, p. 613) that includes the rate constant k of the electron transfer at the individual hopping sites which have mean hopping distance δ.

$$D_E = D + k\delta^2 C_p^* b \tag{8.13}$$

The bulk concentration of the hopping sites is C_p^* and b is a numerical constant whose value depends on the dimensionality of the diffusion (2-D or 3-D). The modulation of the diffusion in conducting polymer can then be affected by the degree of swelling of the polymer matrix caused, for example, by organic vapors or by some unspecified morphological changes of the layer. In some cases, this process is sufficiently reversible to qualify as a possible mechanism for selective sensing of organic solvent vapors.

8.1.1.3 Surface Resistance

Chemical modulation of the surface conductivity is the principle of operation of some of the most commercially successful chemical sensors, the high temperature semiconducting oxide sensors. They are known by their brand name "Figaro sensors." They are discussed in detail in Section 8.2.2.1. The reason for their commercial success lies in the fact that their performance and cost match exactly the specific practical needs of many applications, particularly those of the automotive industry. They have been described in great detail, from the point of view of both the underlying physics and chemistry (Morrison, 1994; Logothetis, 1987).
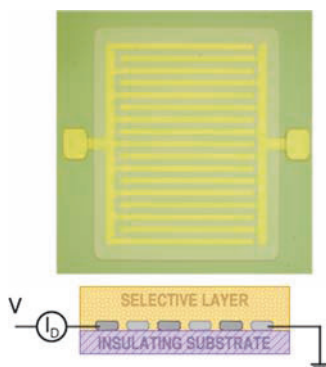
Fig. 8.4 Interdigitated electrodes used in planar chemiresistors

8.2 Chemiresistors

The distinguishing features of these sensors are the simplicity of their preparation and simple instrumentation. Typically, they are implemented as "interdigitated electrodes" (Fig. 8.4), which are patterned on an insulating substrate, and the signal is at zero frequency, that is, dc. The purpose of "interdigitation" is to increase the contact area. The selective layer is then deposited by solvent casting, evaporation sublimation, and so on, over the pattern of these electrodes. In most cases, the contact area for the positive and the negative terminal is the same. This fact does not aid the interpretation of the results.

8.2.1 Thin Metal Film Sensors

One of the simplest, albeit least discussed, chemiresistors is the mercury vapor sensor. The heart of this device is a thin (7–40 nm) film of gold evaporated on a ceramic or glass substrate. At such thicknesses, the films are continuous, with the resistivity between 300 and 1,500 Ωcm. The sensors are usually connected as a pair in a DC bridge configuration with one sensor acting as the reference.

When a DC voltage is applied to the electrodes, the conduction electrons travel in the thin film between the two electrodes. Their mean free path is comparable to the thickness of the film. When foreign species are adsorbed at the metal film surface, they act as a scattering center for the electrons, thus effectively increasing the resistance of the surface layer (Fig. 8.5). This is equivalent to impurity (or carrier) scattering observed in the bulk of doped semiconductors except that in this case the scattering centers are formed at the surface by the adsorbate. For thin gold films, the contribution of the surface layer which is affected by adsorption is significant as compared to the overall conductance of the interior.
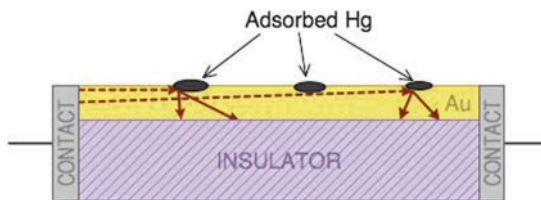
Fig. 8.5 Thin Au film sensor for mercury vapor and other adsorbing species
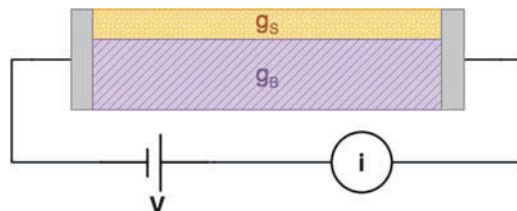
Fig. 8.6 Relationship between surface σ_s and bulk conductivity σ_B in thin-film semiconducting oxide chemiresistors

There are several gaseous species that have a high affinity for the gold surface, such as mercury vapor and halogen–divalent sulfur compounds. The dynamic range depends on the adsorption isotherm and the sensitivity is affected by the thickness of the film. Obviously, for thicker films the contribution of the above effect to the overall change of resistivity is reduced. Although in the original paper (McNerney et al., 1972) the response is described as "reversible," the actual commercial instrument is used in the integrating mode followed by a thermal desorption pulse at 150°C. The dynamic range for mercury is between 1 and 1,000 pg/m^3 and for H$_2$S, 3–500 ppb. The response time is 13 s. There are practically no interferences from water, oxygen, SO$_2$, or hydrocarbons.

8.2.2 Semiconductor Film Sensors

The emphasis on the thickness of the selective layer is linked to the mode of interaction between the analyte and the selective layer. If this interaction takes place exclusively at the surface of the selective layer, then the bulk conductivity does not contribute and represents only a shunt which decreases the signal-to-noise ratio. This is a typical case of chemiresistors based on inorganic materials (Fig. 8.6). On the other hand in chemiresistors based on organic semiconductors, the signal usually originates in the bulk of the selective layer. In that case, the response time of the sensor is affected by its thickness.

8.2.2.1 Inorganic Semiconductor Chemiresistors

The principle of operation of these sensors can be summarized as follows. There is a finite density of electron donors (e.g., adsorbed hydrogen) and/or electron acceptors (e.g., adsorbed oxygen) that are bound to the surface of a wide bandgap semiconducting oxide, such as SnO_2 or ZnO. They represent surface states that can exchange electrons with the interior of the semiconductor, thus forming a space charge layer which lies close to the surface (Fig. 8.7). The position of the surface state relative to the Fermi level E_F of the semiconductor depends on its affinity to electrons. If its affinity is low, it will donate electrons (reducing agent) to the space charge region. If the gas is an electron acceptor (i.e., oxidizing agent), it will be positioned above the Fermi level and it will extract electrons from the space charge region. By changing the surface concentration of the donors–acceptors, the conductivity of the space charge region is modulated. In an n-type semiconductor, such as SnO_2 and ZnO, the majority carriers are electrons, therefore the change of the surface conductivity $\Delta\sigma_S$ from (8.12) is

$$\Delta\sigma_S = e\mu_s\Delta n_s \tag{8.12}$$

where μ_s is the electron mobility at the surface. The excess density of charge carriers (Δn_s) in the space charge region of thickness d is obtained by integrating the difference of the electron density in the space charge region and in the bulk n_b over the thickness d.

$$\Delta n_s = \int_0^d (n_z - n_b)dz \tag{8.13}$$

The change of the surface conductance ΔG_s is then

$$\Delta G_s = \Delta\sigma_S \frac{W}{L} \tag{8.14}$$

For an n-type semiconductor (e.g., ZnO), an increase of the surface concentration of the electron donor (e.g., hydrogen) will increase the conductivity and vice versa. On the other hand, adsorption of hydrogen on p-type oxide (e.g., CoO) will decrease its conductivity.

The bulk conductance (G_b), which is not modulated by the surface reactions, is shown as a parallel conductor (Fig. 8.7).

$$G_b = n_b e\mu_b \frac{Wd}{L} \tag{8.15}$$

Subscript b refers to the bulk quantities and d is the total thickness of the oxide layer. Because the overall conductance is measured, it is advantageous to have the films as thin as possible. The relative change of conductance of the whole device is obtained from (8.12), (8.14), and (8.15) and by assuming that $\mu_b \sim \mu_s$.

$$\frac{\Delta G_s}{G} = \frac{\Delta n_s}{n_b d} \tag{8.16}$$
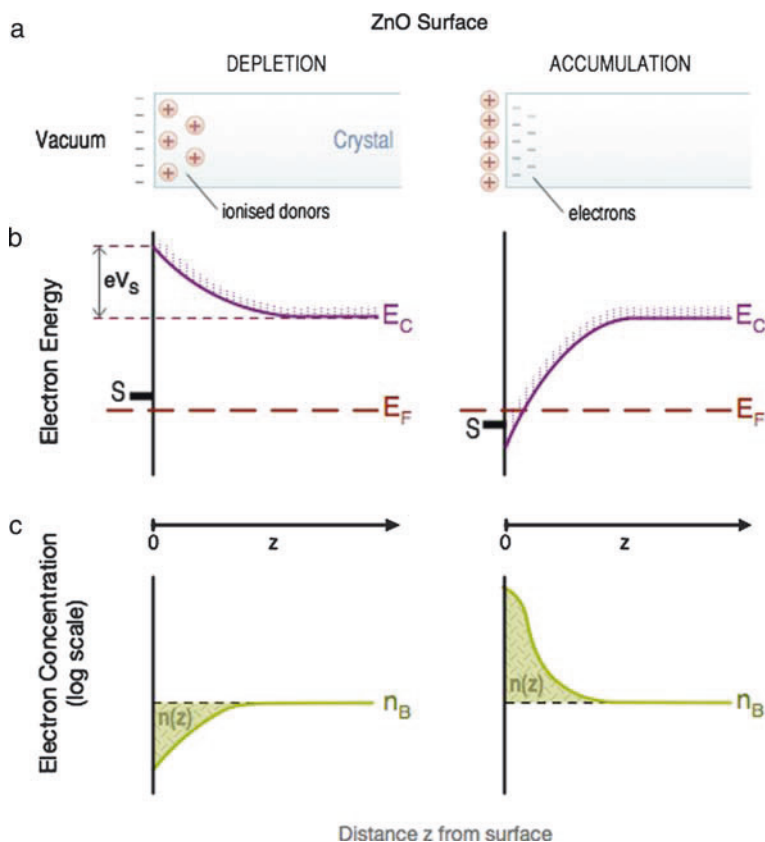
Fig. 8.7 Modulation of space charge by adsorbing analyte at the surface of ZnO (adapted from Heiland, 1982)

For a high value of the relative conductance change (high sensitivity), it is necessary to have a low density of bulk carriers and a thin film. The space charge thickness is typically 10–100 nm, depending on the doping.

As an example, let us consider a typical value of the excess surface state density $n_s = 10^{12}$ electrons cm^{-2} and the bulk density $n_b = 10^{17}$ electrons cm^{-3}. This means that for $100\,\mu m$ thick film, the sensitivity $\Delta G_s/G_b = 10^{-3}$. However, for $d = 10\,nm$ the sensitivity is 10 and for thin films ($<100\,nm$) the space charge region extends throughout the whole film thickness.

It is convenient to discuss the elemental aspects of the sensing mechanism assuming a regular crystal lattice, that is, a monocrystal. Indeed, the fundamental studies of the reactions at oxide surfaces have been done on such well-defined surfaces. On the other hand, practical sensors are prepared by techniques that yield polycrystalline or amorphous layers. For the dimensions of the film at which the surface conductance begins to dominate the overall conductance, the morphology of film becomes the
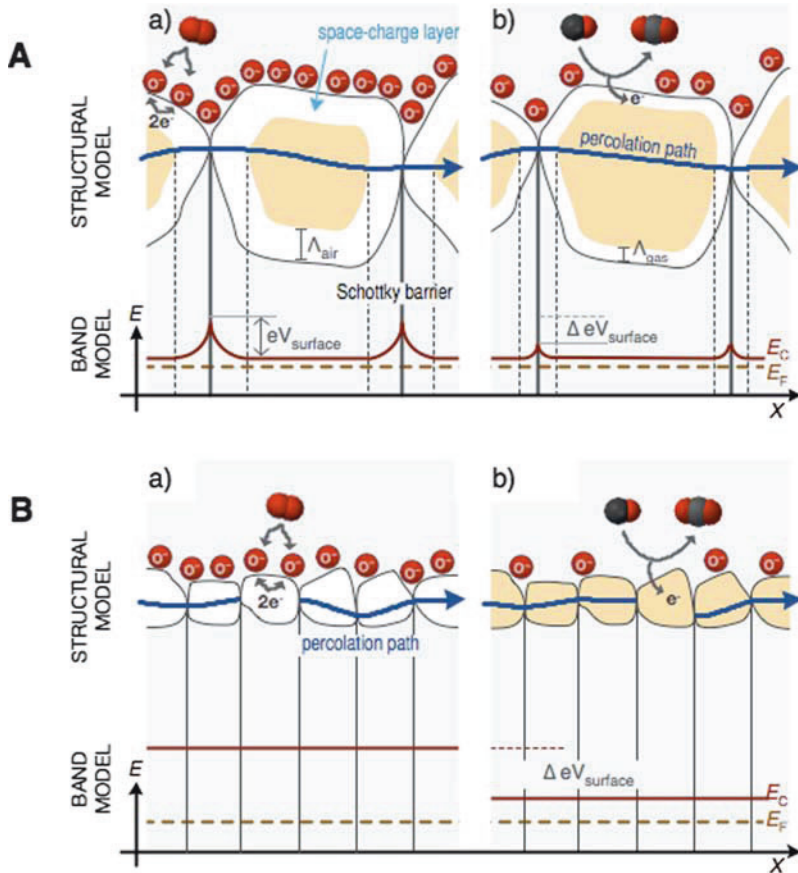
Fig. 8.8 Model of charge transport and band structure of macroscopic polycrystalline (A) and (B) "nano" crystalline metal oxide semiconductors (**a**) in their initial state and (**b**) after exposure to reducing gas (adapted from Franke et al., 2006)

most important parameter. The schematics of the two types of the oxide surface are shown in Fig. 8.8.

There have been several proposed mechanisms for the operation of these sensors (Göpel, 1985; Franke et al., 2006). They all seem to converge on the existence and modulation of the Schottky barrier heterojunctions formed between the grains of the polycrystalline layer. They are equivalent to a chain of resistive elements connected in series. The density of surface states affects the depth of the Schottky barrier and depends on the interaction with the adsorbate (Fig. 8.8). The size of the grains apparently plays a major role. As the diameter of the grains decreases to below 5 nm, the space charge is smeared and the relative response of the sensor increases (Fig. 8.9).

Although any adsorbate that is capable of being involved in the electron exchange with the surface can act as the surface state and dominate the space charge, oxygen and hydrogen are by far the most ubiquitous species. The chemical interactions that
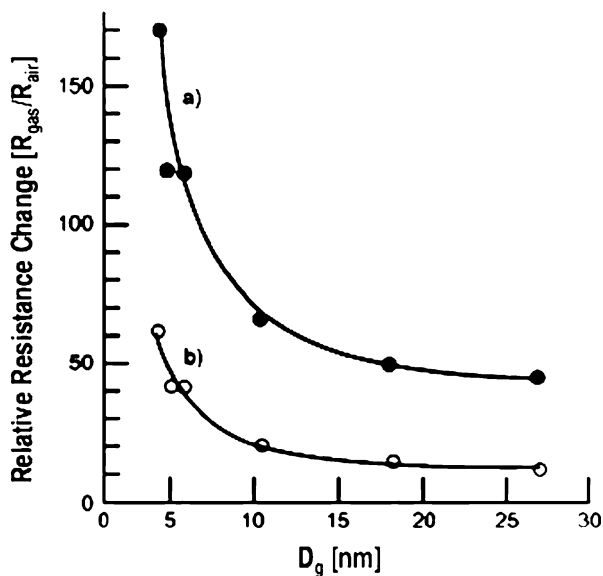
Fig. 8.9 Effect of grain size (D_g) on the response of the SnO$_2$ sensor to (**a**) 800 ppm of H$_2$ and (**b**) 800 ppm of CO in air, at 573 K (adapted from Frank et al., 2006)

lead to the changes of density of surface states are complex and are part of the surface catalytic processes. In addition to the above-mentioned interaction of the adsorbates, the diffusion of lattice defects from the bulk of the crystal also takes place. The defects can act as donors or acceptors. Oxygen vacancies V_O are presumed to be intrinsic donors (Heiland, 1982) whereas adsorbed O$_2$ and O species act as electron acceptors.

$$(Zn^{2+}O^{2-})_{\text{lat}} = V_{O^+} + e^- + \frac{1}{2}O_2\,(\text{gas})$$

and

$$(Zn^{2+}O^{2-})_{\text{lat}} = V_{O^{2+}} + 2e^- + \frac{1}{2}O_2\,(\text{gas})$$
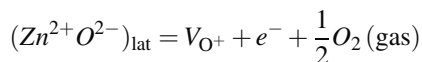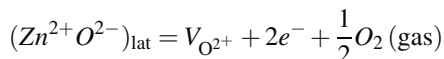
Their relative population at the oxide surface to a large extent determines the type of chemical interactions that take place. The issue of selectivity of these surfaces has been treated comprehensively and in great detail elsewhere (Göpel, 1985; Morrison, 1994). Here we mention only a few examples.

Reactions of the crystal surface involving molecular hydrogen are important not only from the point of view of hydrogen sensing but namely for all reactions in which the abstraction of the hydrogen atom is part of the catalytic mechanism. At low temperatures, adsorbed hydrogen is the source of additional electrons in the space charge.

$$H + (ZnO)_{\text{lat}} = (ZnOH^+)_{\text{lat}} + e^-$$
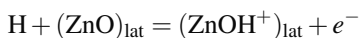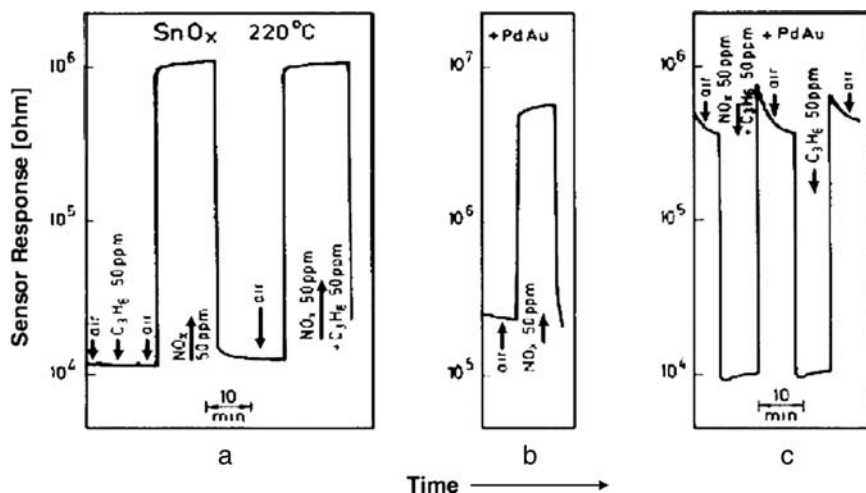
Fig. 8.10 (**a**) Response of a thin-film ($\sim$100 nm) SnO_2 sensor to NO_x and ethane; (**b**), (**c**) response of the SnO_2 sensor modified with a layer of Pd/Au to the same gases (adapted from Heiland, 1982)

At higher temperatures, the lattice itself becomes reduced.

$$2H + (ZnO)_{lat} = H_2O + Zn(0)$$
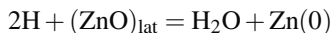
Another ubiquitous species, water, is also involved in the surface redox chemistry of SnO_2.

$$H_2O + (SnO_2)_{lat} = (HO - SnO^{2+})_{lat} + (OH^-)_{lat} + e^-$$
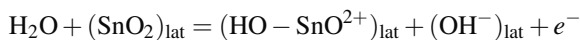
The oxidation and degradation of small hydrocarbon molecules have many similarities (Fig. 8.10). These reactions have been studied extensively by a variety of surface techniques such as Thermal Desorption (TDS), ion mass spectroscopy, specular infrared spectroscopy, Secondary Ion Mass Spectroscopy (SIMS), X-ray Photoelectron Spectroscopy (XPS), and so on. The reaction pathways and the final products depend on the type of oxygen-bearing species, which in turn depends on the doping and morphology of the oxide layer. This is the major reason why results obtained with different oxide sensors in different laboratories do not always agree.

As an example of the complexity of the surface chemistry involved, let us consider oxidation of methane on a (101) SnO_2 surface.

$$CH_4(g) = CH_3(s) + H(s)$$

Here symbols g and s refer to the phase (gas or solid) in which the species are bound, rather the phase of the species itself. At higher concentrations, two adjacent methyl groups combine and react with the lattice oxygen.

$$2CH_3 + O_{lat} = CH_3CH_2O(s) + H(s)$$
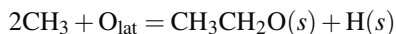
The surface-bound ethoxide ion is dehydrated and desorbs from the surface as ethylene, leaving behind an oxygen vacancy.

$$CH_3 CH_2 O(s) + H(s) = H_2O(g) + C_2H_4(g) + V_O$$
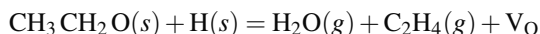
An alternative pathway may lead to formation of surface-bound acetate by the second oxidation step involving lattice oxygen.

$$CH_3 CH_2 O(s) + O_{lat} = CH_3 COO(s) + C_2 H_4(g) + 2H(s)$$
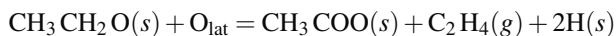
Had the first oxidative coupling not taken place, the final product would have been adsorbed formate.

$$CH_3 + 2O_{lat} = HCOO(s) + 2H(s)$$

All the above species have been detected in various quantities at oxide surfaces. The discussion of this example serves mainly to show that catalytic reactions at oxide surfaces are very complex. This is a mixed blessing from the sensing point of view. It provides a broad spectrum of reactions that could be used. On the other hand, it can lead to great variation in the results obtained with only slightly different sensors. Another drawback of such a complex and diverse mechanism is the relatively slow time response which, in most cases, is limited by the rates of the chemical reactions (Fig. 8.10). Naturally, one tendency of the current research in this field is to increase the selectivity of the surface reactions by introducing additional catalytic control, for example, by incorporation of catalytic metals, metal clusters, and other surface modifiers.

When elevated temperature is needed for its operation, an integral heater can be incorporated in the design of the sensor (Fig. 8.11). It is possible to measure either the steady-state value of the resistance or, for slower reactions, the initial slope as the function of concentration. The principal gases for which this type of sensor has found most widespread application are lower hydrocarbons. However, CO, SO_2, H_2, NO_x, and other gases have been measured. Table 8.1 lists the sensitivity of a commercial SnO_2 sensor to some of these gases.

8.2.2.2 Organic Semiconductor Chemiresistors

Most types of organic chemiresistors are constructed on a nearly identical principle. A thin ($0.01–1\,\mu m$) organic chemically selective layer is applied over the interdigitated electrode array (Fig. 8.12). This film is deposited on an insulating substrate (quartz, glass, alumina, etc.) by evaporation (sublimation), as a Langmuir–Blodgett film or by solvent spin-casting. Electrode spacing is typically $100–500\,\mu m$ and the total electrode area is typically $0.1–1\,cm^2$. The overall resistivity of these films is on the order of $10^8–10^{11}\Omega cm$. For a typical applied voltage 1–5 V, the currents are in the nA range. A variety of electrode materials has been used, ranging from the transition metals (e.g., Ni, Al, Ti, Cu, etc.), which are covered with surface oxide, to oxide-free noble metals (Au, Pt, Ag; see Section 8.1.1.1).
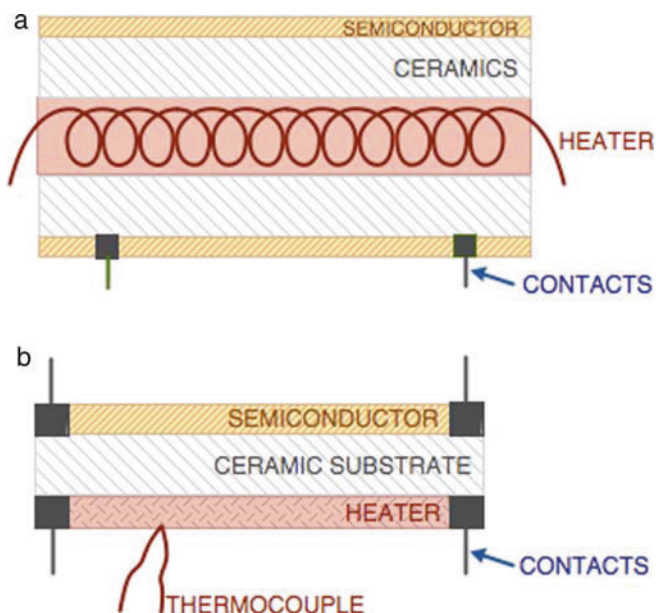
Fig. 8.11 Two versions of a SnO_2 sensor with incorporated heating element: (**a**) a ceramic tube sensor and (**b**) planar sensor (adapted from Heiland, 1982)

Table 8.1 Response of commercial SnO_2 sensor to organic vapors in air

Gas	Formula	Concentration in Air (ppm)
Hydrogen	H_2	1,000
Methane	CH_4	20,000
Ethylene	C_2H_6	4,000
Ethane	C_2H_6	20,000
Ethanol	C_2H_5OH	10,000
Propane	C_3H_8	4,000
Butane	C_4H_{10}	20,000
Carbon dioxide	CO_2	40,000
Carbon monoxide	CO	4,400
Methyl mercapten	CH_3SH	11
Sulfur dioxide	SO_2	10,000

The adhesion of noble metals to the insulating substrates is generally poor. In order to overcome this problem, a thin layer (10–50 nm) of "adhesive" metal (Cr, Ti, W) is deposited under the noble metal. Often this "adhesive" metal migrates through the noble metal and forms an oxide layer that significantly alters its surface
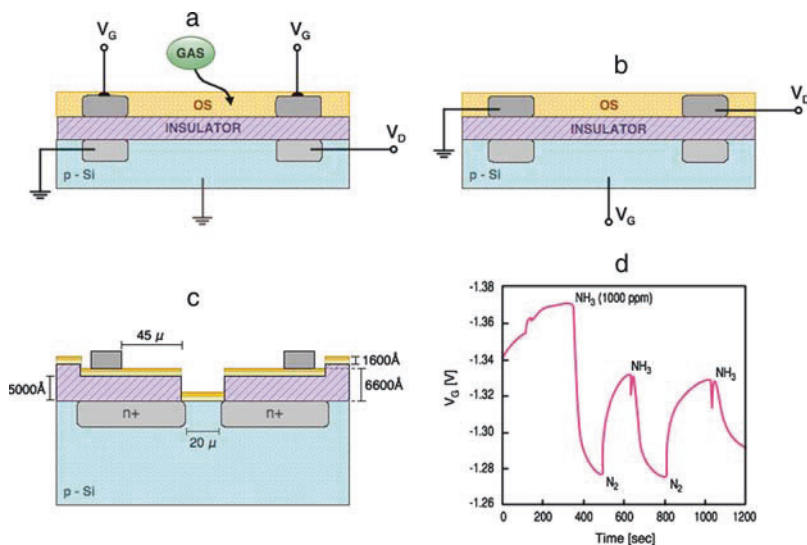
Fig. 8.12 Dual purpose (**a**) CHEMFET, (**b**) OFET structure used for chemical sensing. (**c**) Schematic of the device and (**d**) response to ammonia

properties. This is a general problem that affects all planar structures using thin films of noble metals (Josowicz et al., 1988). The Schottky barrier-type behavior has been observed for chemiresistors utilizing a metal/phthalocyanine (1000 Å)/metal combination (Nguyen and Potje-Kamloth, 1999) that responds to NH_3. Unfortunately, because the electrodes are of comparable size, it is not possible to identify the origin of the response (Diaz et al., 1995). The emphasis on the "solubility" is motivated by the ease of application. In principle, other thin-film deposition techniques can be used, such as sublimation or even electrochemical growth. The latter is somewhat problematic, inasmuch as the supporting substrate must be insulating and the electropolymerized material must bridge the gap between the adjacent electrodes (Potje-Kamloth, 2002).

The transduction mechanism is similar to that discussed already in Section 6.2.3.3, dealing with work function modulation. Both change of mobility and density of carriers can be used. In the latter, the changes can be induced by interaction of an organic semiconductor with a donor–acceptor gas (Janata, 1991), resulting in modulation of concentration of available carriers. Such a process can be viewed as secondary doping (MacDiarmid and Epstein, 1995). A typical electron–acceptor gas is NO_2, which changes the lead phthalocyanine chemiresistor conductivity (Nguyen and Potje-Kamloth, 1999). A thin phthalocyanine layer is deposited over the interdigitated grid of Pt electrodes. At 150°C the device responds reversibly between 1 ppb to 10 ppm of NO_2. Because the response is independent of the thickness of the film, it has been concluded that the response is due to adsorption. However, the thickness insensitivity could also be due to the modulation

of the contact resistance. This seems to be the case with chemiresistors utilizing Langmuir–Blodgett films of structurally similar porphyrins and Cu interdigitated electrodes. An example of electron–donor gas interaction is the ammonia sensor based on polyaniline (Domansky et al., 1998). It again shows a broad dynamic range and reversible response.

Thin-layer organic sensors operating on the principle of changes of bulk conductivity caused by swelling have been used extensively in so called "electronic noses" (Albert et al., 2000). In that case, the polymer matrix is inert (i.e., insulating) and the charge transport is facilitated by electron hopping between embedded conducting particles, such as graphite or metal powder. Because they lack defined electronic structure, these materials are not "semiconductors," but rather "poor conductors."

Organic Field-Effect Transistors (OFETs)

When a third electrode is added to the structure shown in Fig. 8.12a, below the substrate, it creates an electric field in the insulating substrate, which is perpendicular to the path of the current between the two contacts to the selective layer. Such a structure became known as the "Organic Field-Effect Transistor" (OFET). It is discussed here and not in Chapter 6 because it belongs to the family of organic chemiresistors.

The first devices of this type appeared in the early 1970s (Barbe and Westgate, 1970). The field of OFETs literally exploded in the mid-1980s when the first claims of flexible "polymer electronics" were made (Kozeuka et al., 1987). The subject was first reviewed by Horowitz (1998) and hundreds of OFET-related papers have been published since.

It has been known that operation of OFET is affected by the chemical ambient, namely by the presence of oxygen and humidity. This "problem" has been conveniently turned into an "advantage" by promoting OFET as new type of "chemical sensor" for gases and vapors (Torsi et al., 2000). They lack an explicit quantitative relationship between the concentration of the detected species and the device output. The results are invariably given as "change of current" or "relative change of current."

The problem with the use of OFETs as chemical sensors has its origin in the physics of OFETs (Chen et al., 2007). All resistors in the path of the current (Fig. 8.1a) can contribute to the sensor signal. Moreover, because of the coplanar geometry all these resistors can be affected the "gate electric field." The origin of response in CHEMFET can be unambiguously attributed to the modulation of work function of the gate organic semiconductor because the current passes through the silicon (Fig. 8.12a). On the other hand, the response of OFET to chemical stimulus is largely undefined. This ambiguity can be attributed to the different modes of operation of the two "field-effect structures" shown in Fig. 8.12.

Introduction of the donor–acceptor gas can result in the net change of the carriers in the OS, leading to both change of WF and conductivity of the layer (Fig. 8.12b). Furthermore, at low voltages the height of the Schottky barrier at the contact can be modulated as the result of changing the WF difference between the contact metal and OS. If the guest molecule is, for example, an electron acceptor, it can also change the charge-transfer resistance at the drain contact by acting as a redox species at that junction. Finally, the gas interaction resulting in the change of the concentration of carriers changes the bulk conductivity of the OS. Any or all of these mechanisms can operate simultaneously, making a rational assignment of the origin of the OFET sensor response impossible. A sustained passage of current through the OS also induces chemical changes at the contact interface with the electrode. It is a consequence of Faraday's law, which applies equally to electrolysis in all materials.

8.3 AC Conductometric Sensors

Frequency as an experimental variable offers additional design flexibility. This approach has several advantages. The most important one is the lack of polarization of the contacts. The second one is the fact that equivalent electrical circuit analysis can be used that aids in elucidation of the transduction mechanisms. Perhaps the most important distinguishing feature of this class of conductometric sensors is the fact that their impedance is measured in the direction normal to their surface. In fact, there may be no requirement on their DC conductivity and their response can be obtained from their "capacitive" behavior. In the following section, we examine so-called "impedance sensors" (or "impedimetric" sensors; see Fig. 8.1b).

8.3.1 Gas Membrane Sensors

The idea of separating the gas sample by a gas-permeable membrane from the actual internal sensing element is common to several types of electrochemical and some optical sensors. The potentiometric Severinghaus electrode and the amperometric oxygen Clark electrode have already been discussed. Actually, most types of sensors can be used in this configuration and the conductometric sensor is not an exception (Bruckenstein and Symanski, 1986).

The theory of operation of the conductometric gas membrane sensor has been experimentally verified in detail for CO_2 and SO_2, and sensors for H_2S and NH_3 based on the same principle have also been made. The basic transport and equilibration processes are the same as in the Severinghaus electrode (Section 6.2.2). Upon entering the aqueous solution inside the cell, the gas dissociates to its constituent ions. Because each dissociated species contributes to the overall conductivity, the specific conductance Λ of the cell is

$$\Lambda = 10^{-3}(\lambda_{H^+}C_{H^+} + \lambda_{X^-}C_{X^-}) \tag{8.17}$$

where λs are the equivalent ionic conductances and X^- is the anion. The composition of the internal solution does not vary significantly, therefore it is legitimate to use concentrations instead of activities. If the cell initially contains pure water, the specific conductivity is described by (6.76) and (8.17).

$$\Lambda = 10^{-3}(\lambda_{H^+} + \lambda_{X^-})KP_{HX}^{1/2} \tag{8.18}$$

The experimental conductance G_{cell} depends on the cell geometry as characterized by the cell constant κ. Therefore, the output of the sensor is related to the partial pressure of the gas as in (8.19).

$$G_{\text{cell}} = \frac{10^{-3}[(\lambda_{H^+} + \lambda_{X^-})(KP_{HX})^{1/2}}{\kappa} \tag{8.19}$$

The magnitude of the dissociation constant K plays an important role in the response characteristics of the sensor. For a weakly dissociated gas (e.g., CO_2, $K = 4.4 \times 10^{-7}$), the sensor can reach its equilibrium value in less than 100 s and no accumulation of CO_2 takes place in the interior layer. On the other hand, SO_2, which is a much stronger acid ($K = 1.3 \times 10^{-2}$), accumulates inside the sensor and its repsonse time is in minutes. The detection limit and sensitivity of the conductometric gas sensors also depend on the value of the dissociation constant, on the solubility of the gas in the internal filling solution, and, to some extent, on the equivalent ionic conductances of the ions involved. Although an aqueous filling solution has been used in all conductometric gas sensors described to date, it is possible, in principle, to use any liquid for that purpose. The choice of the dielectric constant and solubility would then provide additional experimental parameters that could be optimized in order to obtain higher selectivity and/or a lower detection limit.

8.3.2 Dielectrometric Sensors

When a periodically changing excitation signal is chosen for the operation of chemiresistors, they can be used to detect changes of capacitance (Fig. 8.1b). Therefore, the dielectrometric sensors rely on the chemical modulation of one or more equivalent circuit capacitors, either through the change of the dielectric constant of the chemically sensitive layer or through the chemical modulation of the interfacial charge.

Because of its high dielectric constant (78.5 at 25°C), water is the primary (but not the only) species that can be measured by following the changes of the values of the capacitors in the equivalent circuit. Therefore, this section could be also called "hygrometry," because humidity sensors are the largest represented group in this category. However, because "DC hygrometers" (water-sensitive chemiresistors)
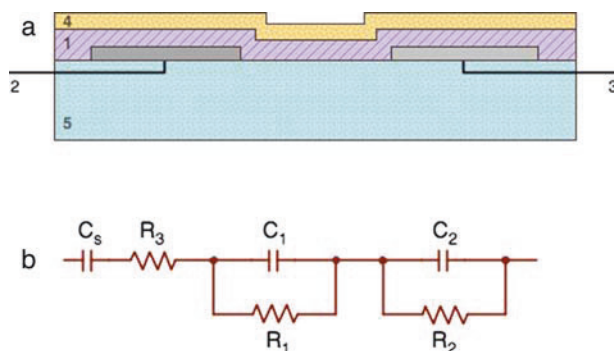
Fig. 8.13 A capacitive "impedance" sensor. Schematic diagram (**a**) and equivalent electrical circuit (**b**) 1-vapor absorbing layer; 2-Cr/Ni/Au plate of the capacitor (C1); 3-Ta plate (C2); 4-top, porous metal plate; 5 insulating substrate

have also been constructed and because dielectrometers for other molecules can also be made, we stay with the general term dielectrometry.

The simplest dielectrometer is a capacitor containing a layer of a material that can more or less reversibly take up chemical vapor of a given dielectric constant. The diagram of such a sensor and its equivalent electrical circuit is shown in Fig. 8.13 (Garverick and Senturia, 1982).

There are several things that need to be pointed out on this design. First, the upper electrode is porous, either because it is very thin (10–20 nm) or because it is deposited under such conditions that it "cracks." In any case, the polymer beneath it comes into contact with the gas or vapor. Obviously, it is difficult to make robust electrical connections to the top electrode. Fortunately, it is not necessary because it forms an electrically floating plate, which is common to the two capacitors: one between the Cr, Ni, Au plate (C_1) and the other between the top and the Ta plate (C_2). The corresponding leakage resistances are R_1 and R_2. The response of this sensor to water vapor is shown in Fig. 8.14.

8.3.3 Interfacial Charge Sensors

It was pointed out Chapter 5 that interfaces always separate charges. Direct measurement of charge, that is, by field-effect transistor, although possible in principle, proved to be difficult due to the nonideal properties of the capacitive (i.e., "ideally polarized") interfaces (Section 5.1). However, it is possible to evaluate interfacial charge from the impedance analysis of electrodes. Electrochemical Impedance Spectroscopy (EIS) is a very powerful technique for the study of complex electrochemical processes (Gabrielli, 1995; Park and Yoo, 2003). It has also been used for detection of selective binding of small metal ions to monolayers of immobilized ligands (Gafni et al., 1996). Sensing of small ions by this technique may
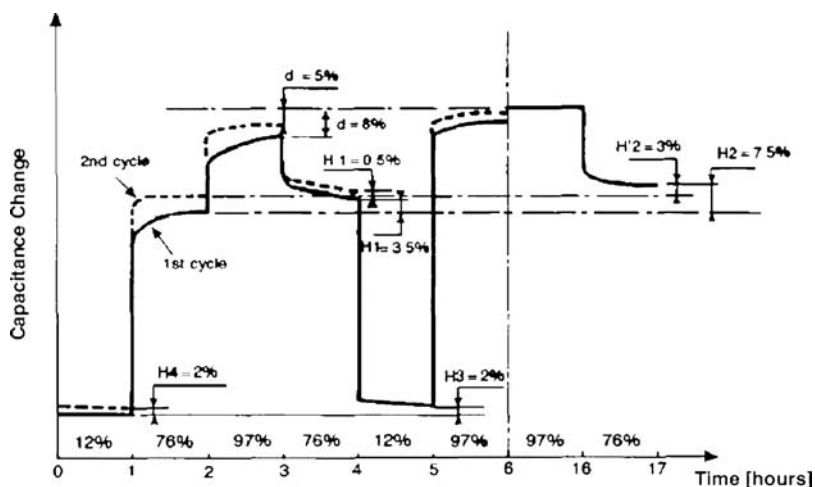
Fig. 8.14 Response to water vapor of hygrometer shown in Fig. 8.13

not be the most practical, however, this possibility has prompted development of "impedimetric" biosensors (Berney, 2004). The basic idea is to immobilize a suitable probe biomolecule, such as antibody, ssDNA, and so on, at the surface of an electrode. Such molecules are invariably polyelectrolytes, which means that they carry a net charge which dominates the charge at the solution side of the interface. When the complementary target molecule selectively binds to the immobilized probe molecule, the resulting adduct has a different charge which is then evaluated from the complex plane impedance analysis (i.e., the Nyquist plots, Chapter 5). Such measurement is done by imposing a small amplitude AC voltage and measuring the corresponding current at a given frequency. Small, mobile redox ions are added to the solution.

These impedimetric biosensors suffer from the "virtually irreversible" binding discussed in Chapter 2. This is the main reason why they do not qualify as direct biosensors. They have been used in conjunction with the ferro/ferricyanide redox couple as the indicator of the blocking of the surface (Radi et al., 2005). Unfortunately, the binding event is so strong that the analysis has to be run in an assay format. The result of such a procedure is shown in Fig. 8.2.

Even more serious are the interferences resulting from the nonspecific adsorption of charged species (small ions and polyelectrolytes) at the electrode surface. Also, the nonspecific effect of the ionic strength changes the value of the diffuse part of the double layer, again resulting in changes of the overall capacitance of the electrode. Invariably, the results are reported for rigorously controlled experimental conditions. It is fair to say that EIS is a powerful and indispensable technique for investigation of electrochemical interfaces, but due to its vulnerability to many interferences, it may not be robust enough to be a practical sensing principle in most applications.

Food for Thought #8

Bubbles

1. Why are bubbles of gas formed at the working electrode, but not at the auxiliary electrode? After all, the gas is evolved at the same rate at both electrodes, according to (8.4) and (8.5).

Closed Circuit

2. In high-frequency conductometric titrations, the electrodes are placed on the outside wall of the glass titration vessel. Does this arrangement violate the golden rule requiring a "closed electrical circuit" in electrochemical experiments? What requirement does it place on the applied frequency?

Capacitive pH Sensors

There are many physical parameters that change with pH of the sample: surface charge density, absorbance, polymer swelling, and so on.

3. What happens when conductivity changes with pH?
4. Is it possible to obtain the information about the pH of the solution without a reference electrode using "capacitive" measurement?
5. What makes the glass electrode/reference electrode combination so special?

Channel Conductivity Sensors

6. It has been proposed to use self-assembled or Langmuir–Blodgett bilayer membranes with embedded selective receptors as sensors for corresponding agonists. This concept mimics, for example, gated neuronal membrane function. The information is obtained from chemically modulated changes of the overall impedance of the membrane measured in the normal direction. The barrier membrane is usually represented in a "cartoon" as a row of ampiphilic molecules standing at 90° to the substrate. In reality, the molecules are always tilted. Consider what effect this tilting has on the background value of impedance measured in the direction normal to the surface.

Immunocapacitor

It has been proposed to use a Si/SiO_2/electrolyte structure for the detection of immunochemical reaction. The antibodies are immobilized at the surface of the SiO_2 and their interaction with the antigen is monitored by observing the shift of the inflex point on the C–V curve, using 150 mV p–p, 1 KHz modulation.

7. Suggest an equivalent circuit diagram corresponding to this experiment and explain what causes the "shift" of the inflex point.

Ion Channel Sensors

Blocking and unblocking of the access of a redox marker to the metal electrode surface is the core idea of "ion channel sensors." The selectivity of such a "gated" sensor hinges on the highly specific binding of the agonist molecule to the receptor that then opens or closes the channel. The blocking action itself can be dependent on either charge or size. Demonstration of such a sensing principle has been performed successfully in laboratory settings with strictly controlled "samples." Try to answer the following questions.

8. Is this really a reversible sensor?
9. What would be the most likely interferences encountered by such a sensor in a real sample?

Polyelectrolyte Sensors

Dendrimers are large, highly branched macromolecules that can be decorated with various charged and functional groups within their interior and also on the periphery. Thus, den-64-[Ru(tpy)2]$^{2+}$ carries 128 positive charges yet, at ionic strength >0.1 M, it adsorbs according to the Langmuir isotherm.

10. What does this say about the electrostatics of the adsorption process?
11. Based on this fact, what would you say about the prospect of detecting other large polyelectrolytes based on their charge?
12. What type of sensor would you select that would utilize the unusually large binding capacity of dendrimers?

Missing Warburg

A gold electrode with self-assembled monolayers, in the absence of any redox species, is an almost ideally polarized (i.e., purely capacitive) interface. A Nyquist

plot for a pure capacitor should be a straight vertical line. Yet, EIS of such an electrode shows a semicircle and even a straight line (i.e., not at 45°), resembling Warburg impedance.

13. Comment on this apparent discrepancy between the capacitive model and reported shapes of the real Nyquist plots.

Symbols

A	Cross-section
A_R	Richardson constant for thermionic emission of electrons
C	Capacitance
C_P^*	Concentration
D	Diffusion coefficient
D_g	Grain size
d	Thickness
E	Electric field
$E(\omega)$	Voltage
E_F	Fermi level
e	Charge of electron
G	Electrical conductance
$G(\omega)$	Frequency-dependent conductance
h	Planck's constant (6.625×10^{-27} erg s)
$I(\omega)$	Current
J	Junction current
j	Current density
K	Dissociation constant
k	Heterogeneous rate constant
L	Length
m	Mass
n	Number of moles
n_s	Density of charge carries in the space charge region
n_b	Density of charge carriers in the bulk
R	Resistance
T	Temperature
V	Applied voltage
W	Width
Y	Admittances
$Z(\omega)$	Impedance
δ	Hopping distance
Λ	Specific conductance
ϕ_B	Average barrier height
κ	Cell constant

λ	Ionic conductance
μ	Mobility
σ	Conductivity
ρ	Resistivity

References

Albert, K.J., Lewis, N.S., Schauer, C.L., Sotzing, G.A., Stitzel, S.S., Vaid, T.P., and Walt, D.R. (2000) *Chem. Rev*. 100, 2595–2626.

Barbe, D.F. and Westgate C.R. (1970) *J. Phys. Chem. Solids* 31, 2679.

Bard, A.J. and Faulkner, L.R. (2001) *Electrochemical Methods, Fundamentals and Applications.* Wiley.

Berney, H. (2004) Impedometric and amperometric chemical and biological sensors. In: V.M. Mirsky (Ed.) *Ultrathin Electrochemical Chemo- and Biosensors Part II*. Springer.

Bruckenstein, S. and Symanski, J.S. (1986) *J. Chem. Soc. Farad. Trans*. I, 82, 1105.

Diaz, A.F., Nguyen, My T., and Leclerc, M. (1995) Electronically conducting soluble polymers. In: I. Rubinstein (Ed.) *Physical Electrochemistry*. Marcel Dekker.

Domansky, K., Baldwin, D.L., Grate, J.W., Hall, Jing Li, T.B., Josowicz, M., and Janata, J. (1998) *Anal. Chem*. 70, 473–481.

Franke, M.E., Koplin, T.J., and Simon, U. (2006) *Small* 2, 36.

Gabrielli, C. (1995) Electrochemical impedance spectroscopy; Principles, instrumentation and applications. In: I. Rubinstein (Ed.) *Physical Electrochemistry*. Marcel Dekker.

Gafni,Y., Weizman, H., Libman, J., Shanzer, A., and Rubinstein, I. (1996) *Chem. Eur. J*. 2, 759.

Garverick, S.L. and Senturia, S.D. (1982) *IEEE Trans. El. Dev*. ED-29, 90.

Göpel, W. (1985) *Prog. Surf. Sci*. 20, 9.

Heiland, G. (1982) *Sens. Act*. 2, 343.

Horowitz, G. (1998) *Adv. Mat*. 10, 365.

Janata, J. (1991) *Anal. Chem*. 63, 2546–2550.

Josowicz, M., Janata, J., and Levy, M.J. (1988) *Electrochem. Soc*. 135, 112.

Kozeuka, H., Tsumura, A., and Ando, T. (1987) *Synth. Met*. 18, 699.

Logothetis, E.M. (1987) Oxygen Sensors for Automotive Applications. In: D.R. Turner (Ed.) *Chemical Sensors*. Electrochemical Society, Pennington, NJ, p. 142.

MacDiarmid, A.G. and Epstein, A.J. (1995) *Synth. Met*. 69, 85–89.

McNerney, J.J., Buseck, P.R., and Hanson, R.C. (1972) *Science*, 178, 611.

Morrison, S.R. (1994) Chemical sensors. In: S.M. Sze (Ed.) *Semiconductor Sensors*. Wiley.

Nguyen, V.C. and Potje-Kamloth, K. (1999) *Thin Solid Films* 338, 142.

Park, S.-M. and Yoo, J.-S. (2003) *Anal. Chem*. 455A.

Potje-Kamloth, K. (2001) Conducting polymer based Schottky and heterojunctions diodes and their sensor application. In: Nalwa, H.S. (Ed.) *Handbook of Surfaces and Interfaces of Materials, Vol. 5*. Academic Press, p. 445–494.

Potje-Kamloth, K. (2002) *Crit. Rev. Anal. Chem*. 32, 121.

Potje-Kamloth, K. (2008) *Chem. Rev*. 108, 367.

Torsi, L., Dodabalapur, A., Sabbatini, L., and Zambonin, P.G. (2000) *Sens. Act. B* 67, 312.

Chapter 9
Optical Sensors

9.1 Introduction

Interaction of electromagnetic radiation with matter occurs over a broad range of frequencies and usually in a highly specific way (Table 9.1). Study and use of these interactions is in the domain of spectroscopy, which provides information ranging from the electronic structure of atoms to dynamics of polymeric chains.

In the most general conventional arrangement, electromagnetic radiation passes through the sample, which may be solid, liquid, or gas. The interaction of this radiation with the sample is evaluated from the change of some optical parameter and is related to the concentration of the analyte (Fig. 9.1). The radiation impinging on the sample can contain all wavelengths or only selected ones (i.e., monochromatic), but the wavelength selection must take place somewhere along the optical path. The absorption of the primary radiation can also be coupled to other, nonoptical, effects, for example, to the increase of temperature or pressure or to the change of electrical conductivity.

The principles of optical sensors are anchored in classical spectroscopy; only the elements of the experiment and their arrangement are different. Generally speaking, in a spectroscopic measurement the sample is placed in a spectrophotometer, in a well-defined path of the light beam, and the emerging radiation is captured by the detector. In an optical sensor, the light is guided out of the spectrophotometer, allowed to interact with the sample, and then reintroduced into the spectrophotometer either in its primary or secondary form for further processing. The necessity of guiding and manipulating the light over a distance dictates the choice of materials and the applications in which the optical sensors can be used. Light conduits are now available for more or less all wavelengths, although they can be expensive. The development of fiberoptic sensors started with the introduction of communication optical fibers and hardware, and it still parallels the developments in the communication field to a large extent. The spectroscopic knowledge serves as the basis for the development and use of optical sensors. In that respect, the situation is similar to

Table 9.1 Molecular spectroscopy by optical sensors

Region	Wavenumber (μm^{-1})	Wavelength (μm)	Transition	Information	Energy (kJ/mol)
Far IR	10^{11}–10^{12}	10–1,000	rotational	Interactions	0.5
IR–Raman	10^{12}–10^{14}	1–10	vibrational	functional groups	4
UV–Vis	10^{14}–10^{16}	0.1–1	electronic	ionization energy	40–400
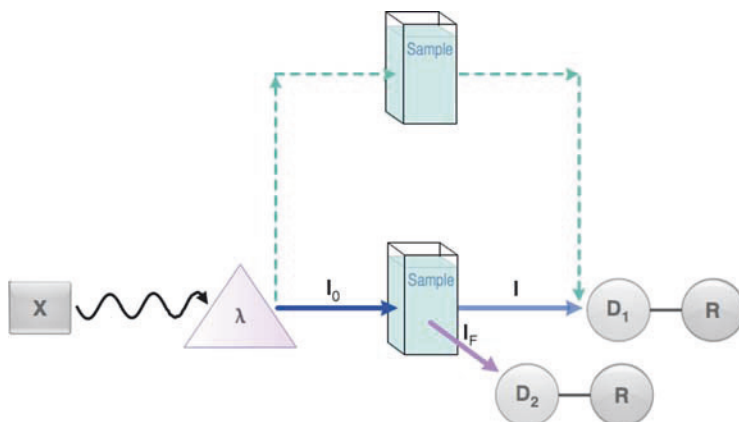
Fig. 9.1 General arrangement of spectroscopic absorption/fluorescence measurement: λ-wavelength selector, X-light source, D-detector, R-recorder. Intensity of initial (I_0), transmitted (I), or fluorescent (I_F) light is related to the concentration of the sample C

that of the electrochemical sensors, the design and operation of which are informed by the knowledge base of general electrochemistry.

The concept of dual character of light is particularly relevant to the discussion of optical sensors. In this introductory section, the basic quantized (corpuscular) aspects of light as they relate to optical sensors are reviewed first, followed by a brief review of physics of optical waveguides and optical fibers which rely on wavelike (continuous) properties of light. Detailed information can be found in analytical (e.g., Skoog et al., 1998) and specialized textbooks (e.g., Hollas, 2004).

9.2 Corpuscular Properties of Light

Light consists of two sinusoidally oscillating, electric **E** and magnetic **H** fields, which are randomly, but perpendicularly, oriented (Fig. 9.2a). They oscillate in phase and move through the optical medium with propagation velocity v. In vacuum light travels with maximum velocity $c = 2.9979 \times 10^8\,\mathrm{m\,s^{-1}}$. The frequency of oscillation ν, wavelength λ, and velocity v are related by

$$\lambda \nu = v \tag{9.1}$$
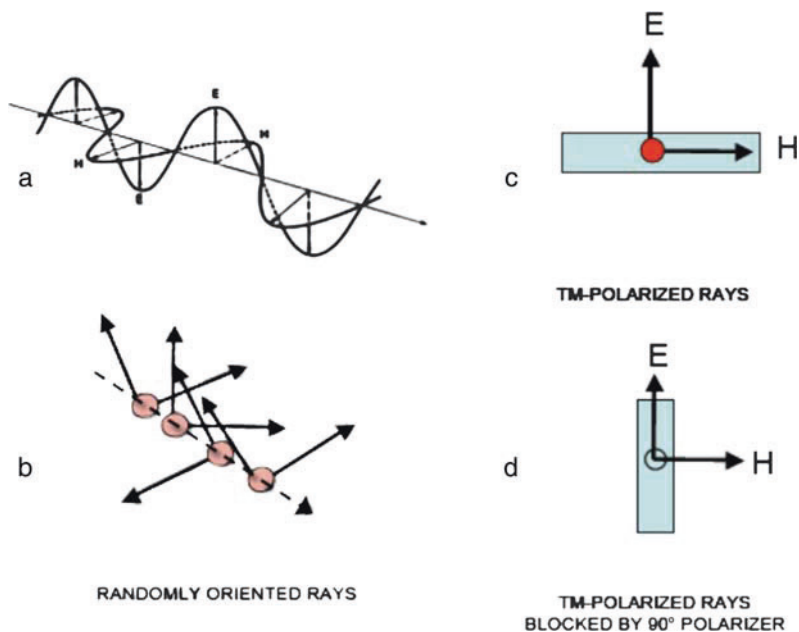
Fig. 9.2 (**a**) Electromagnetic light propagation; (**b**) randomly polarized light; (**c**) polarization of light by passage through polarizing medium; (**d**) blockage of light by cross-oriented polarizer. The intensity of the light is indicated by the color of the red dot

A derived quantity is the wave number $\overline{v}$, which is the reciprocal of λ and specifies the number of wavelengths per unit distance (cm^{-1}). The energy contained in light is quantized. A quantum of energy ΔE_λ is one photon. Mathematically, this is defined as

$$\Delta E_\lambda = hv = \frac{hc}{\lambda} = hc\overline{v} \tag{9.2}$$

where h is Planck's constant $(6.625 \times 10^{-27} ergs)$. From this expression we see that the energy of electromagnetic radiation increases with frequency and with the wave number (see Table 9.1). Equation (9.2) is the foundation of the "corpuscular" aspect of light.

As the wavelength (i.e., frequency) is scanned, the light energy and its degree of absorption by the matter changes. The relationship between the relative absorption and the wavelength is called the absorption spectrum and is characteristic of the absorber. Examination of the energy belonging to the different parts of the spectrum (Table 9.1) shows that these interactions involve different properties of the molecules. The domain of optical sensors is in the ultraviolet, visible, as well as infrared, parts of the spectrum. This means that the interactions can be observed for the transitions at electronic (UV–Vis) as well as the IR vibrational levels (Fig. 9.3). These interactions correspond to the changes of free energy on the order of tens of kilo Joules per mole.
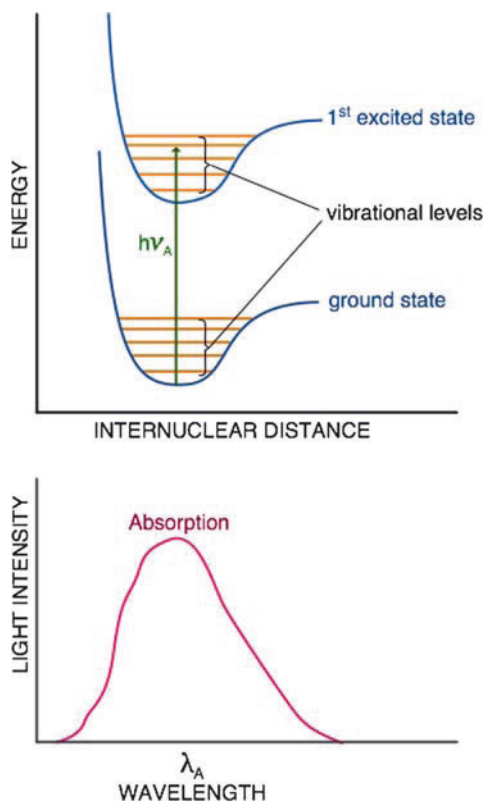
Fig. 9.3 Electronic transitions leading to absorption of light

When a molecule absorbs a photon of an exact quantum of energy it moves from its electronic ground state to its higher electronic excited state. Within each electronic state, the molecule can occupy different vibrational energy levels that are separated from each other by energy differences of less than $4\,\mathrm{kJ\,mol^{-1}}$. (Note: Figure 9.3 is not drawn to scale. The difference between two electronic levels is much larger than the difference between the individual vibrational levels.) The ability of the molecule to absorb light is expressed as molar absorptivity ε_λ, which in the analytical literature is also known as the extinction coefficient.

9.2.1 The Lambert–Beer Law

A step common to most optical sensing processes is the absorption of a photon. The quantitative relationship between the concentration of the absorber C and the attenuation of intensity of incident light is known as the Lambert–Beer law. For its derivation, it is convenient to invoke the "particulate" character of light, which says

that the photon is an "optical particle" which carries a discrete quantum of energy. The intensity of a beam of light of unit cross-sectional area S is

$$I = N \tag{9.3}$$

where N is the number of photons/unit area. When light passes through an absorber of elementary volume $dV = Sdx$, the number of photons absorbed equals dN, resulting in decrease of light intensity.

$$dN = -dI \tag{9.4}$$

The probability P of absorption of a photon is proportional to the number of absorbing molecules present in the elementary volume.

$$P = \frac{dN}{N} = \varepsilon_\lambda C(Sdx) \tag{9.5}$$

where C is molar concentration and the molar absorptivity, ε_λ, is the proportionality constant. Thus, from (9.3) through (9.5)

$$\int_{I_0}^{I} \frac{dI}{I} = -\varepsilon_\lambda C \int_{x=0}^{L} dx \tag{9.6}$$

The intensity of the incident beam is I_0 and I is the final intensity. Integration of (9.6) over the pathlength L and conversion to a base-10 logarithm yields the Lambert–Beer law, which relates absorbance A to concentration.

$$A = -\log I/I_0 = \varepsilon_\lambda CL \tag{9.7}$$

From this equation, we see that the changes of attenuation of the initial beam are equally affected by the changes of the optical path length and/or by the changes of the concentration. In a normal spectrophotometric experiment, the optical path L is constant and defined by the spacing of the transparent cuvette windows. A similarly well-defined relationship often does not exist in optical sensors.

Molar absorptivity is a property of a material or chemical species that can be described mathematically as a function of wavelength. It is an additive property. This means that for samples containing multiple absorbing species within the same optical path and at the same wavelength, the absorbance A is given as

$$A_\lambda = L(\varepsilon_{1,\lambda} C_1 + \varepsilon_{2,\lambda} C_2 + \ldots \varepsilon_{n,\lambda} C_n) \tag{9.8}$$

where Cs are concentrations of the individual species and $\varepsilon_{n,\lambda}$ are the molar absorptivities at given wavelengths λ_v. Because the shapes of the absorption spectra of different species are different (i.e., the εs are different functions of λ) it is possible to perform multiwavelength measurements. Equation (9.8) then serves as a tool of multivariate spectrophotometry of complex mixtures (see Chapter 10).

Let us pause for a moment and review the factors that affect the physical selectivity of optical sensors. First of all, the resonant nature of the interaction between light and absorber is one of the most important factors in achieving selectivity. It is further enhanced by the geometric effects that distinguish the spatial distribution of the fluorescent radiation from the linear character of the primary absorption step. Additional selectivity accrues from the use of polarized light and from the time-resolved fluorescent and phosphorescent measurements. On the other hand, some selectivity is lost due to the line broadening. In summary, the selectivity inherent in the transduction step in optical sensors is far higher than in any other type of chemical sensor. Obviously, as does any other sensor, optical ones also utilize all the benefits of selectivity originating from the interactions in the chemically selective layer.

9.2.2 Luminescence

The excited molecule can lose its energy through dissipation, that is, by rapid weak collisions with its neighboring molecules. In that case, the energy of the molecule falls within the excited state towards its lowest vibrational level. If the molecule remains long enough in the lowest vibrational level of its excited state, it can make the transition to the ground state either by collisions with other molecules (i.e., by thermalization) or radiatively. In the latter case, the energy loss occurs in a quantized fashion and the emitted light again has the characteristic fluorescent emission frequency (Fig. 9.4a). In general, molecules in the gas and liquid phase do not have a fixed orientation with respect to the direction of the excitation radiation. Therefore, the emission of fluorescence occurs in all directions. Because the absorption and emission processes are symmetrical, the corresponding spectra also have a symmetrical appearance.

The absorbing molecule or its specific part is called the chromophore, and the emitting molecule (or its part) is called the fluor. Energy dissipation by strong collisions is clearly a competing process, which limits the efficiency of the fluorescence. Different molecules have different abilities to dissipate energy from the fluor, partly because of their concentration, but also because of their stereospecific position with respect to the excited molecule. Molecules that can dissipate the energy from the excited molecule by a thermal process even at low concentrations are called quenchers. The intensity of emitted light I_F depends on the quantum efficiency ζ, on the concentration of the fluor C_F, on its molar absorptivity ε and on the pathlength L. The quantum efficiency is defined as

$$\zeta = \frac{\text{photons emitted}}{\text{photons absorbed}}$$

Because usually only a fraction of the emitted light is captured by the detector, the collection factor γ is also introduced. The final expression is then

$$I_F = \gamma \zeta \varepsilon C_F L \tag{9.9}$$
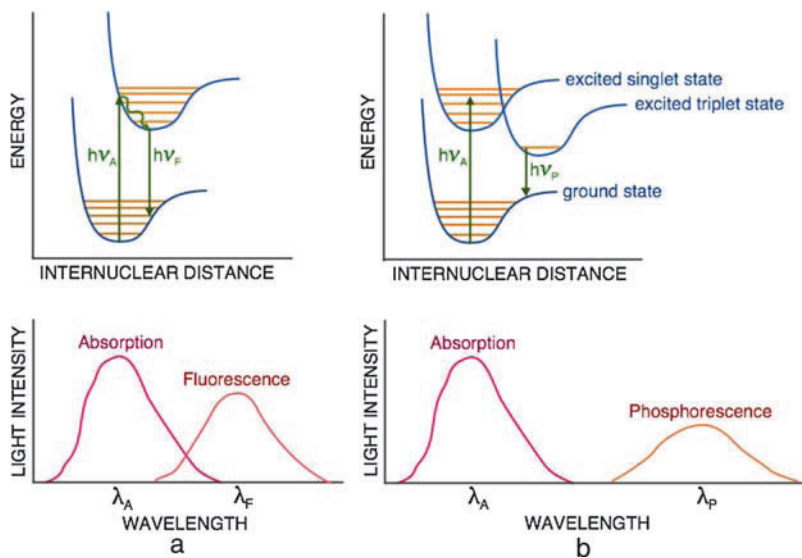
Fig. 9.4 Electronic transitions leading to (**a**) fluorescence emission at λ_F and (**b**) phosphorescence at λ_P

Because the specific and the nonspecific quenching are competitive processes, the fluorescence lifetime ($1/e$ value of the original intensity) is a characteristic parameter that can be used to describe both the concentration of the fluor and the quencher. Typical fluorescence lifetimes are in nanoseconds.

Another effect that could be used for chemical sensing is photobleaching. It is the degradation of the fluor due to the irreversible chemical reactions of the molecule in its excited state. As the process of photobleaching is irreversible, it does not lend itself to reversible chemical sensing.

There is an inherent similarity between the spectrum and an electrochemical current–voltage curve that is important from the point of view of chemical selectivity. In both cases, the x-axis (voltage or wavelength) is directly related to the energy. In electrochemistry, this energy corresponds to the transfer of electrons between the analyte and the electrode. It is related to the standard electrochemical potential. In optical interactions, molar absorptivity is probabilistically related to the excitation energy of the molecule.

There is, however, a fundamental difference that makes optical selectivity based on ε_λ far more important than the electrochemical selectivity derived from the applied potential. In electrochemistry, at a given applied potential, all electrochemical reactions that have equal or lower energy will take place. This means that the selectivity defined by the choice of the electrode potential only is very limited. In contrast, optical transitions are resonant and fall off on both sides of the absorption maximum. If not enough energy is applied, or if too much energy is applied, the transition (and therefore the interaction) is forbidden; that is, its probability is low.

Consequently, many more individual absorption processes can be accommodated on the frequency (energy) axis. Their actual number is indirectly proportional to the line-width. According to (9.2), the quantum of energy associated with the transition that would correspond to a single spectral line is sharply defined. Such a line spectrum is observed, for example, in atomic vapors. On the other hand, spectral lines of more complicated molecules, even in gas phase, are broader. This is due to the fact that the transition between two electronic states is complicated by the presence of multiple vibrational levels within each state. Furthermore, in the condensed phase, these vibrational levels are strongly affected by interactions with the surrounding molecules.

The second reason for line broadening is the fact that the electronic (absorption) transitions may take place between the ground state and the first or higher excited states. In contrast, because fluorescence always takes place between the first excited state and the ground state, the fluorescent spectra are usually sharper than the corresponding absorption spectra.

A third reason for the line broadening lies in the change of the internuclear distance between the atoms during the absorption of the energy quantum by the molecule (the Franck–Condon effect). This effect has no direct implication for chemical sensors except that line broadening in general represents a certain loss of chemical selectivity. Vibrational transitions require much lower energy than electronic transitions. The individual states are less affected by the environment of the molecule and as a consequence the spectral lines are much narrower, even in the condensed state. Because of this, infrared sensors can benefit to a far greater extent from the selectivity based on wavelength (energy) than their UV–Vis counterparts.

Similar to fluorescence, but of a slightly different origin is phosphorescence. The transition to the first excited (singlet) state on absorption of the quantum of light is the same. If there is a triplet state (with two unpaired electrons) lying energetically close to or below the first excited state, a nonradiative singlet–triplet transition can take place (Fig. 9.4b). This transition has relatively low probability because it is "forbidden" by the rule which states that the change of multiplicity of electron spins is "not allowed." The radiative decay from the triplet state to the ground state, analogous to the fluorescent emission, then takes place. Because there are multiple rearrangements that must occur during this process, phosphorescence has much longer lifetime (up to milliseconds) than fluorescence.

Another process mentioned in the context of chemical sensing is chemiluminescence. In this case, the excited state is created by a chemical reaction, which then emits light according to the following scheme.

$$A + B \Rightarrow X* \Rightarrow X + \eta v$$
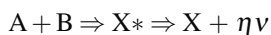
There are also several electrochemical reactions in which the electroluminescence occurs as a result of chemical reactions following the initial electron transfer. If biomolecules are involved, the emission is called bioluminescence. Because the reactions leading to the excited state are invariably irreversible, these schemes are

generally used in chemical assays and in bioassays. Their discussion lies outside the scope of this book.

The electronic transition that fits the definition of (9.2) is generally called an "allowed" transition. When the sample is irradiated with light of a higher frequency (higher energy) than would correspond to any electronic transition, most of the light passes through the sample without attenuation. Only a small fraction is scattered, giving rise to very faint lines corresponding to the transitions within vibrational levels (Fig. 9.5).

These lines belong to the Raman spectrum and are observed at a right angle to the incident beam. These transitions are "not allowed," in the sense of the Bohr model of atom. However, if we realize that, in quantum mechanics, "allowed" means "having a higher probability" and, conversely, "not allowed" means "having a lower probability," then the existence of these transitions can be rationalized. Interaction of light with the molecule in which only the direction of propagation is changed is called Rayleigh elastic scattering. If the molecule absorbs a fraction of energy, weak Raman lines (Fig. 9.6) shifted towards lower frequencies (longer wavelengths) appear. They are called Stokes lines.

Symmetrically with respect to the Rayleigh line lie the somewhat less intense anti-Stokes lines which correspond to the transfer of energy from the molecule to
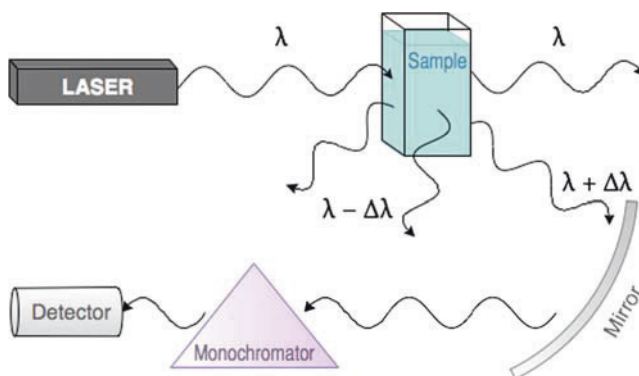
Fig. 9.5 Scattering of incident radiation
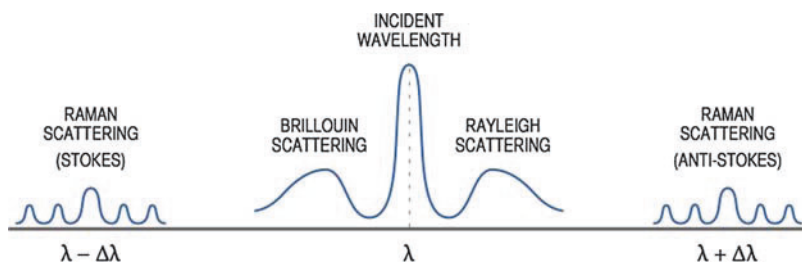
Fig. 9.6 Rayleigh, Raman, and Brillouin scattering lines

the incident light. Closer to the central line are weak Brillouin lines which are caused by the interaction of the light with the phonons in the condensed (i.e., liquid or solid) medium.

From the sensing point of view, it is important that the difference of frequency between the Rayleigh line and one of the Raman lines corresponds directly to the energy difference of the transition between two vibrational levels, according to (9.2). The Raman effect thus provides vibrational information, which is normally in the domain of infrared spectroscopy. Yet, any sufficiently strong source emitting in the visible part of the spectrum, for example, a laser, can be used as the source of the excitation radiation. This is particularly important from the sensing point of view, because the materials for optical fibers and waveguides for transmission in the visible spectrum are readily available, whereas those for the infrared region are still relatively expensive. Another significant advantage of Raman spectroscopy is that it largely circumvents the broad absorption of IR radiation by water.

9.3 Wavelike Properties of Light

When light interacts with matter, several processes take place, sometimes simultaneously. It can be absorbed, which results in a decrease of the primary intensity, or it can be transmitted without attenuation. The propagation velocity v of light depends on the optical density of the medium. It is related to the absolute (vacuum) velocity of light c by

$$n = c/v \tag{9.10}$$

where n is the refractive index of the medium. Because c is the maximum propagation velocity, $n > 1$. Interfaces between two optical materials are particularly important for optical sensing. When a light beam reaches an interface of materials with differing optical densities, its energy (i.e., its color) does not change and its phase must be preserved. Therefore, it must change direction. The resulting phenomena are called reflection and refraction. As we show, these two conditions are essential for guiding the light through optical conduits. If the direction and/or frequency of the light is changed upon interaction with matter, we talk about absorption and scattering. Fluorescence and phosphorescence are special forms of scattering.

When two or more monochromatic light beams interact, the phase relationship between the two beams becomes important. Light can be coherent, in which case the photons of the two beams are in phase and their interaction produces interference fringes with unit visibility. If the interaction is noncoherent, the phase relationship between individual photons is random and the intensities of the two beams add photometrically. The light produced by an optical laser is monochromatic (one energy) and coherent (in phase). A noncoherent light passing through a monochromator (i.e., "wavelength selector") remains noncoherent.

When light passes through a condensed phase consisting of oriented electrical dipoles (polarizer), or when it is reflected from a dipolar interface it becomes polarized (Fig. 9.2). It means that orthogonally randomly oriented **E, M** vectors are now

aligned. If the magnetic vector of polarized light is parallel with the interface and perpendicular to the direction of propagation the light is said to be TM polarized (Transverse Magnetic vector polarized). If the electric vector is parallel with the interface the light is TE polarized (Transverse Electric vector polarized). Polarized light can be attenuated by the second polarizer rotated at angle ϑ with respect to the primary polarizer. At $\vartheta = 90°$ the transmission of polarized light is completely blocked.

A typical optical sensor consists of modules found in conventional spectroscopy, for example, source, monochromator, cuvet, detector, and so on. The big difference is that light travels between these modules through optical fibers or waveguides. As it travels, the beam of light interacts with interfaces between media of different optical densities and therefore of different refractive indices. The laws that govern its behavior at these boundaries are the most important aspects of the operation of optical sensors (Okamoto, 2000).

When light traveling through the optically dense medium of refractive index n_1 arrives at its boundary with an optically rarer medium of refractive index n_2 (i.e., $n_1 > n_2$), it can be either reflected back to the phase 1 or be refracted (i.e., transmitted through the interface) to the phase 2. Which of these two conditions applies depends on the angle of incidence Θ_1 at the interface (Fig. 9.7). The speed of light is lower in the medium of higher optical density. The phase of the transmitted light and its frequency must be preserved during passage through the interface. Because they are fixed to the wavelength by (9.1), the angle Θ_2 of the transmitted light must change in order to satisfy the condition of phase matching. This is why objects submerged in water appear to be "broken," but have the same color as the original object. Naturally, the angle of reflected light Θ_2 is unchanged and equals the angle of the incident beam.

The geometrical relationships for refracted light are given by Snell's law (Fig. 9.7).

$$n_1 \sin \Theta_1 = n_2 \sin \Theta_2 \qquad (9.11)$$

The angle of refracted light Θ_2 increases faster than the angle of the incident light. When Θ_1 reaches a certain value, the condition of phase matching can no longer be satisfied and all the light is transmitted along the interface (i.e., $\Theta_2 = 90°$). The angle at which this happens is called the critical angle Θ_C. At that angle, $\sin \Theta_2 = 90° = 1$
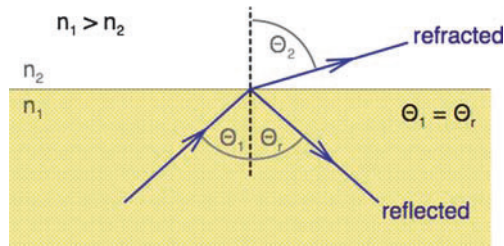
Fig. 9.7 Reflection and refraction of light at the interface of two media of different refractive indices

and (9.11) reduces to

$$\sin \Theta_C = \frac{n_2}{n_1} \tag{9.12}$$

For the air/water interface at 25°C, the critical angle Θ_C is 43.75°. For values of $\Theta_1 > \Theta_C$, all the light is reflected to the optically dense phase. This condition is called total internal reflection. It is the key condition of optical wave guiding and thus the most fundamental condition of optical sensing.

9.3.1 Guiding of Light

In order to propagate the electromagnetic wave along the interface, or to reflect it totally to the optically denser medium, there must be a zero net flux of energy across the interface. However, there is a finite decaying electric field across the interface that extends some distance z into the rarer medium. The amplitude of this evanescent field is given by

$$E_2 = E_0 \exp\left[-\frac{i2\pi n_2}{\lambda}(x\sin\Theta_2 + z\cos\Theta_2)\right] \tag{9.13}$$

At the critical angle of the incident light $\Theta_1 = \Theta_C$ and $\Theta_2 = 90°$. It means that sin $\Theta_2 = 0$ and cos $\Theta_2 = 1$. The electric field penetrating in the perpendicular direction z is

$$E_2 = E_0 \exp\frac{-i2\pi n_2 z}{\lambda} \tag{9.14}$$

Because $n_1 > n_2$, for values of Θ_1 greater than Θ_C, the combination of (9.11) and (9.12) produces the seemingly paradoxical result: "$\sin\Theta_2 > 1$". This result can be accepted because the electric vector is a complex number and because an electromagnetic wave does not "propagate" in medium n_1 but only "penetrates" into it. The standing wave at the interface then forms an evanescent field at the n_2 side of the interface.

The depth of the evanescent field can be determined as follows. From (9.11) and using relationship $\sin^2\Theta_2 + \cos^2\Theta_2 = 1$ we obtain

$$\cos^2\Theta_2 = \left[1 - \left(\frac{n_1}{n_2}\right)^2\sin^2\Theta_1\right] \tag{9.15}$$

For $\Theta_1 > \Theta_C$ the term in the brackets is negative. Therefore,

$$\cos\Theta_2 = i\left[\left(\frac{n_1}{n_2}\right)^2\sin^2\Theta_1 - 1\right]^{1/2} \tag{9.16}$$

Combination of (9.11), (9.13), and (9.16) yields the value of the electric field at the point of reflection at the arbitrary angle of incidence $\Theta_1 > \Theta_C$ and its decay in the

optically rarer medium (n_2).

$$E_2 = E_0 \exp \left\{ -\frac{2\pi i x}{\lambda} \left(n_1 \sin \Theta_1 + i n_2 \left[\left(\frac{n_1}{n_2} \right)^2 \sin^2 \Theta_1 - 1 \right]^{1/2} \right) \right\} \quad (9.17)$$

Separation of two terms in the exponent argument and multiplication by $(i)^2 = -1$ yields

$$E_2 = E_0 \exp \left(-\frac{i 2\pi x}{\lambda} n_1 \sin \Theta_1 \right) \exp \left(\frac{2\pi z}{\lambda} n_1^2 \sin^2 \Theta_1 - n_2^2 \right)^{1/2} \quad (9.18)$$

The first (imaginary) exponential term in (9.18) is responsible for the lateral shift of the reflected wave in the direction x of its propagation (Fig. 9.8).

The second (real) term accounts for the exponential decay of the electric field intensity in the direction normal to the interface. The reflected beam combines with the incident beam, forming a standing electromagnetic wave at the interface (Fig. 9.9). The electric field that penetrates to the optically rarer medium of refractive index n_1, the evanescent field, plays a critical role in many optical sensors based on the waveguiding principle. Its depth of penetration d_p is defined as the distance at which the initial intensity E_0 decays to 1/e of its value. Thus from (9.18), d_p is
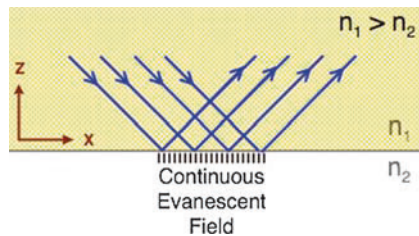
Fig. 9.8 Origin of the continuous evanescent field at the interface of the waveguide
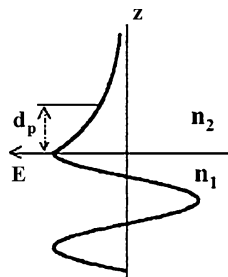
Fig. 9.9 (**a**) Standing electromagnetic wave at the interface and the exponential decay of the intensity of the evanescent field to the optically rarer medium

$$d_p = \frac{\lambda}{2\pi \left(n_1^2 \sin^2 \Theta_1 - n_2^2\right)^{1/2}} \tag{9.19}$$

The depth of penetration plays a critical role in the definition of the "optical volume" that is being sampled. It depends on several parameters, above all on the wavelength of the incident beam, on the angle of incidence, and on the refractive indices of the two media. This consideration is important in sensing applications in which the interaction of a chemical species with this field gives rise to an analytical signal. For common materials and for visible light, d_p is typically between 100 and 200 nm.

Looking at the phenomenon of optical absorption by the medium from the viewpoint of classical wave mechanics, we see that the attenuation of electromagnetic radiation can be attributed to the interaction of the oscillating electric vector with the medium. Any phenomenon involving periodic oscillations can be decomposed to real and imaginary components. Thus, the ordinary refractive index n is the real part of the index of refraction $\tilde{n}$, which can be written as

$$\overline{n_\lambda} = n - ik_\lambda \tag{9.20}$$

where ik_λ is the imaginary part related to molar absorptivity as

$$k_\lambda = \frac{\lambda \varepsilon_\lambda}{4\pi} \tag{9.21}$$

Therefore, in the complex plane of the evanescent field at penetration depth d_p, the Lambert–Beer law is expressed as

$$\log \frac{I}{I_0} = A = -\frac{4\pi k_\lambda d_p C L}{\lambda} \tag{9.22}$$

Substituting for d_p from (9.19) yields

$$A = \frac{2k_\lambda C L}{\left(n_1^2 \sin^2 \Theta_1 - n_2^2\right)^{1/2}} \tag{9.23}$$

where L is the length of the evanescent field along the waveguide (x-direction). Unlike in normal spectrophotometry, the absorbance A is a nonlinear function of the angle of incidence, the refractive indices of the materials of the waveguide, and of the wavelength. Therefore, the optical path in the evanescent field experiments is very complex and dependent on the number of experimental parameters. This is the price that has to be paid for "taking the light out of the spectrophotometer." In chemical sensing, these parameters must be tightly controlled in order to obtain a valid functional relationship between the absorbance and the concentration of the analyte.
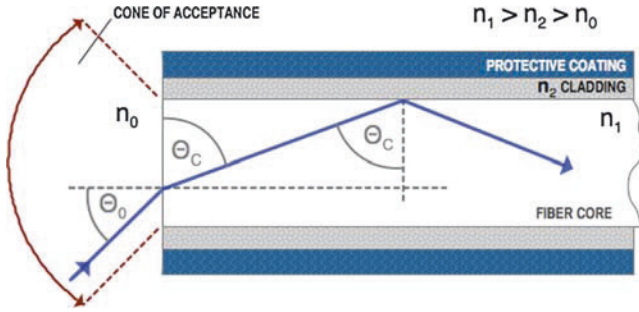
Fig. 9.10 Construction of optical fiber and the coupling geometry defining the cone of acceptance

9.3.2 Coupling of Light

Construction of the optical fiber is shown in Fig. 9.10. The optical core of refractive index n_1 (a typical material is silica, which has $n_1 = 1.6$) is surrounded by optical insulating material called cladding, which has a lower refractive index n_2 (a typical value is 1.5). The outside coating on the fiber is there for mechanical protection and has no special optical function. In order to be coupled in some controllable way, the light beam must enter the fiber within a certain cone whose angle is related to the critical angle for the total reflection at the core/cladding interface. Likewise, the light leaving the distal end of the fiber spreads only over the same solid angle. For the rational design of an optical sensor, it is important to know the shape of this cone.

In the following paragraph we find the angle Θ_0, which defines this cone of acceptance. The primary consideration is the propagation of light within the fiber, for which the critical angle at the core/cladding interface is defined by (9.12). The refraction at the interface between the end of the fiber and the outside medium (n_0) is governed by the Snell's law (9.11)

$$n_0 \sin \Theta_0 = n_1 \sin(90 - \Theta_c) \tag{9.24}$$

Using the relationship $\sin(\alpha - \beta) = \sin \alpha \cos \beta - \cos \alpha \sin \beta$, we obtain

$$n_0 \sin \Theta_0 = n_1 \cos \Theta_c \tag{9.25}$$

With the help of relationship $\cos^2 \Theta_c + \sin^2 \Theta_c = 1$ we obtain

$$n_1^2 \cos^2 \Theta_c + n_1^2 \sin^2 \Theta_c = n_1^2 \tag{9.26}$$

The total reflection of light at the core/cladding interface is

$$n_1^2 \sin^2 \Theta_1 = n_2^2 \tag{9.27}$$

From (9.25) through (9.27)

$$n_1^2 \sin^2 \Theta_0 = n_1^2 - n_2^2 \tag{9.28}$$

or

$$N.A. = \frac{(n_1^2 - n_2^2)^{1/2}}{n_0} \tag{9.29}$$

The $N.A.$ is the numerical aperture of the fiber and defines the cone of acceptance in terms of refractive indices only. The "half-angle" Θ_0 of the "cone of acceptance" is given by (9.28).

9.3.3 Guided Modes

The phenomenon of light reflection has two important implications for chemical sensing. It confines the guided radiation in the optical phase and as such is the basis of transmission of light through waveguides and fibers. The second implication is the existence of the evanescent field on the outside of the waveguide, which represents a sampling space in which the chemical interactions can take place.

When a single beam of light is introduced into a waveguide surrounded by a homogeneous medium of lower refractive index, light bounces from the opposite interfaces and propagates in a zig-zag fashion through the planar waveguide (Fig. 9.11a), or as a spiral through the cylindrical optical fiber (Fig. 9.11b). Naturally, at each point of reflection there is an evanescent field. If there is no interaction within this field and/or within the waveguide, the light will pass without attenuation. If the thickness of the waveguide is much greater than the wavelength of light, the boundary conditions at the two reflecting planes can be satisfied independently of each other.

If we now decrease the thickness of the waveguide and "squeeze" the light, the points of reflection move closer to each other. As they begin to overlap, the reflected beams begin to interfere. This produces another set of conditions that must be satisfied both from the point of view of the single reflection and from the point of view of the constructive interference. As the result of this additional constraint, a finite number of reflection angles exist for each wavelength at which the conditions allow propagation by multiple total reflection. These constructive interferences are called the guided modes of the waveguide (Fig 9.12).
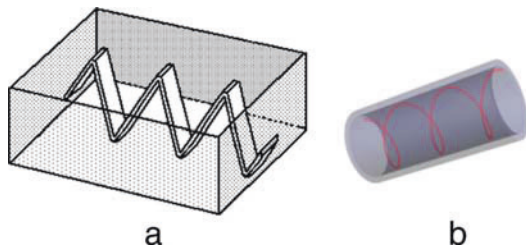
Fig. 9.11 Propagation of light through (**a**) a planar waveguide and (**b**) through a cylindrical optical fiber
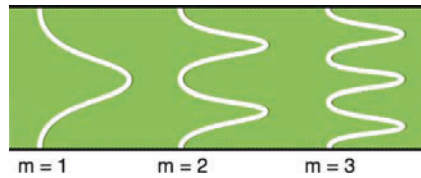
Fig. 9.12 Guided modes in the fiber
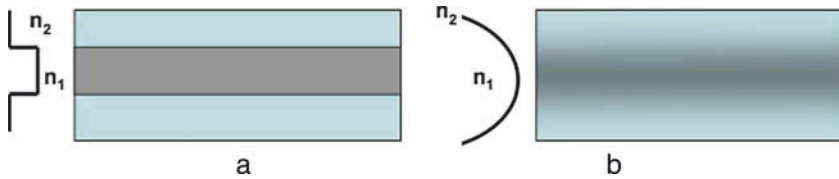
Fig. 9.13 (**a**) Step refractive index and (**b**) gradual refractive index (GRIN) fibers. The line to the left of the fiber represents the value of the refractive index

Obviously, the thinner the waveguide is and the smaller the difference between the refractive indices, the fewer the modes. When the waveguide operates in a multiple reflection mode, light no longer exists as an individual beam but rather appears as a continuous streak and correspondingly continuous evanescent field (Fig. 9.8). This may serve in sensing applications as a large "active area" in which the sensing interaction can take place.

An essentially identical situation exists for the multiple reflection in optical fibers. In this case, the boundary conditions for the multiple reflections and the considerations governing the existence and number of guided modes are formulated in cylindrical coordinates. For a noncoherent source, the light propagated by a fiber again appears as a continuum of intensity. The number of modes M in the fiber of radius r (both r and λ are in μm) is characterized by the "V-number," which is a practical figure of merit for the fiber.

$$V = \frac{2\pi r}{\lambda}(n_1^2 - n_2^2)^{1/2} \tag{9.30}$$

For a fiber in which the interface forms a sharp step between the media of refractive indices n_1 and n_2, the number of supported modes is $M = V^2/2$. For fibers with a Graduated Refractive INdex (GRIN) it is $M \sim V^2/4$ (Fig. 9.13). Thus, for a "step" cylindrical fiber

$$M = \frac{\pi^2 r^2 (n_1^2 - n_2^2)}{\lambda^2} \tag{9.31}$$

Thus, a 1 μm diameter fiber would support one mode of green ($\lambda = 500$ nm) light. Obviously, optical throughput (energy) of such a fiber is very low. A typical fiber of $d < 10$ μm supports thousands of modes.

Flat optical waveguides offer the possibility of increasing the optical throughput but maintaining the monomode internal reflection. This is achieved in the
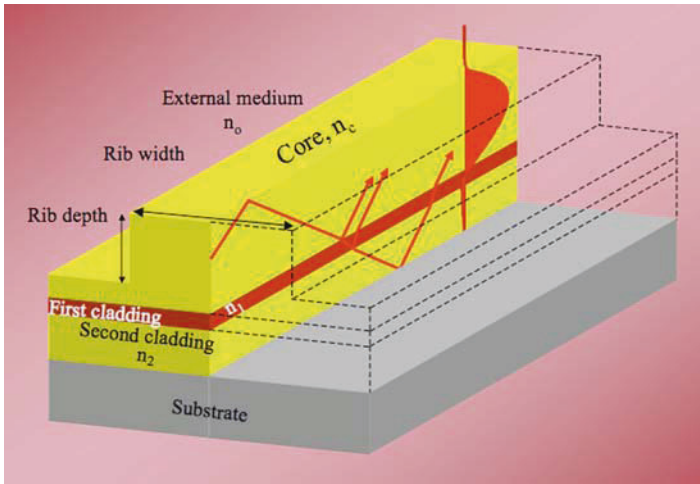
Fig. 9.14 Antiresonant reflecting optical waveguide (adapted from Lechuga et al., 2004)

AntiResonant Reflecting Optical Waveguide (ARROW), in which the single mode is supported inside the 1–4 µm core and the higher modes are "leaked" out through the antiresonant reflection interface layer inside the cladding (Fig. 9.14; Lechuga et al., 2004). Obviously, flat waveguides are not suitable for transmission over long distances, but they can be coupled, for that purpose, with transmission by optical fibers.

9.3.4 Reflectance

Analytical information can be obtained from the measurement of intensity of the reflected light. It carries information about the chemical species at the interface and also about the surface morphology of that interface. When the angle of the incident beam unequivocally determines the angle of the reflected beam we talk about specular reflectance. It occurs at optically smooth surfaces and interfaces, where the amount of scatter is minimal. On the other hand, rough and granular surfaces scatter light over a wide solid angle, due to the interaction between light and the randomly oriented facets of the surface. Chemical and morphological information about the reflector is still obtainable, but the relationship between the attenuation and concentration is less explicit.

9.3.4.1 Diffuse Reflectance

Various models have been proposed that seek, with varying degrees of success, to quantify the reflectance and the concentration. The best known is the Kubelka–Munk model (Kubelka and Munk, 1931) which is sometimes called the
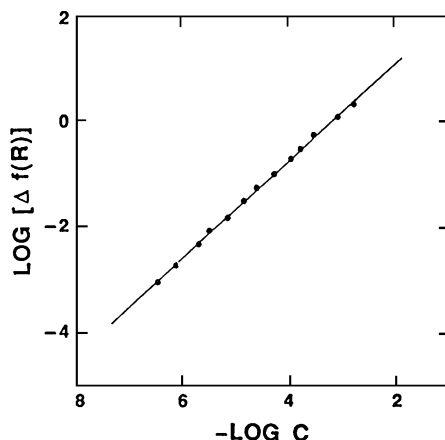
Fig. 9.15 Diffuse reflectance from solid mixture of $KMnO_4$ and $KClO_4$. The solid line is calculated from (9.32) (adapted from Kubelka et al., 1931)

"Lambert–Beer law of reflectance spectroscopy." Function $f(R)$ relates the intensity of the reflected light with absorption (εC) and scattering (S_r) constants.

$$f(R) = \frac{(1-R)^2}{2R} = \frac{\varepsilon_\lambda C}{S_r} \tag{9.32}$$

This relationship works well for many different types of scatterers and for low concentrations of the absorber. The best results have been obtained for mixtures of absorbing and nonabsorbing powders (Fig. 9.15) and for absorbers adsorbed on a colorless solid. A fair agreement has been obtained also for evaluation of colored zones on paper chromatograms. A better result for this type of sample has been obtained using the modified formula (Hecht, 1983)

$$\frac{R_0}{R} = \frac{(1+\beta)^2}{1+\frac{\beta}{Q_r}} \tag{9.33}$$

In this case, the value of reflectance for a pure scatterer R_0 is used to normalize the measurement. Parameters β and Q_r are defined in terms of empirical parameters related to reflection (r), transmission (t), and absorption (μ)

$$\beta = \frac{\mu}{r+t} \tag{9.34}$$

$$Q_r = 1 + \frac{t}{r+t} \tag{9.35}$$

Thus, for a highly reflecting and minimally transmitting sample, $Q_r = 1$, and (9.33) simplifies to

$$\frac{R_0}{R} = 1 + \beta \tag{9.36}$$

The empirical parameters r, t, and μ are used to adjust the equation to the morphology of the sample.

9.3.4.2 Plasmon Resonance

When the interface between a thin ($\sim$50 nm) film of free-electron metal (e.g., Au, Ag, or Al) and a glass prism is irradiated with a TM polarized coherent light, correlated motion of free electrons in the metal leads to the formation of oscillating charge at the metal/glass interface (Fig. 9.16)

$$k_{||} = \frac{2\pi n_g \sin\Theta}{\lambda} \tag{9.37}$$

where n_g is the refractive index of glass, λ is the wavelength of the monochromatic light, and θ is the angle of incidence. A corresponding oscillating charge also appears at the other interface between metal and the sample, which has dielectric constant ε_d.

$$k_{SP} = \frac{\omega}{c}\sqrt{\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d}} \tag{9.38}$$

Collectively the two oscillating vectors are known as surface plasmon (Homola, 2004, 2008), which moves along the surface, but does not radiate energy. Its intensity decays exponentially to the dielectric (sample) in the direction perpendicular to the interface. In that respect, it is similar to the evanescent field.

The dielectric constant is a complex number, having a real and imaginary component. Surface plasmons exist for the frequencies at which the real part of the dielectric constant of the metal ε_m is negative and the imaginary part is greater than the imaginary part of the dielectric constant of the adjacent dielectric ε_d (here the sample). The relationship between the frequency and the value of the propagating vector k_{SP} for the plasmon at the glass/metal surface is given by the resonant condition. It occurs when the velocities of the two vectors $k_{||}$ and k_{SP} are matched.
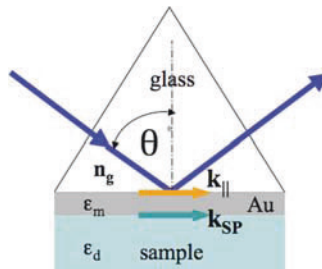
Fig. 9.16 Formation of a surface plasmon at the metal/sample interface
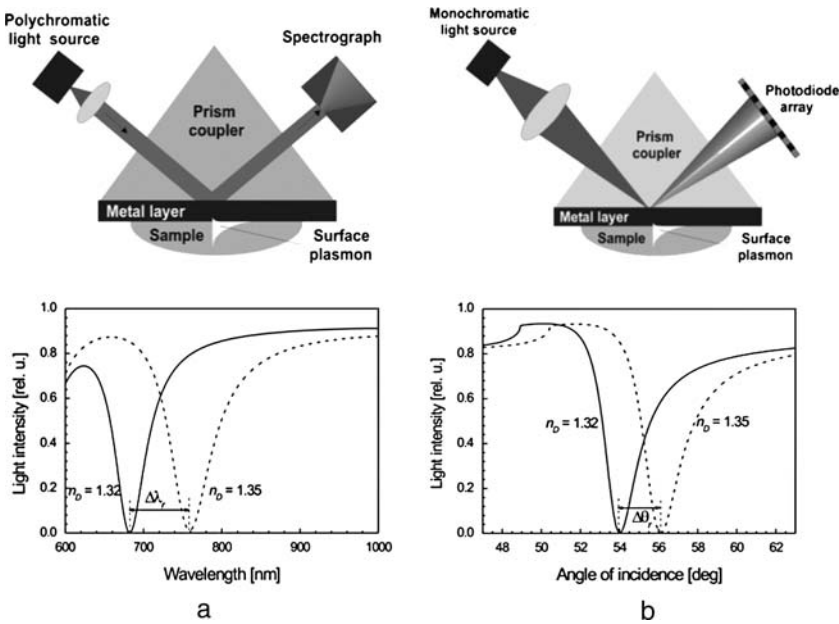
Fig. 9.17 Two modes of operation of SPR sensor: (**a**) fixed angle and variable wavelength and (**b**) variable angle of incidence of monochromatic light

From (9.37) and (9.38), we have for the resonant angle Θ_r

$$\frac{2\pi n_g}{\lambda} \sin \Theta_r = \text{Re}\{k_{SP}\} \tag{9.39}$$

where $\text{Re}\{k_{SP}\}$ is the real part of the surface plasmon. The resonance is seen as the attenuation of intensity of the reflected light as a function of the wavelength of the incident radiation (Fig. 9.17a). The resonant condition can be achieved also by varying the angle of incidence (Fig. 9.17b; Homola, 2004).

Thus the value of the dielectric constant at the sample/metal interface determines the shift of the resonance. When adsorption of molecules at the metal surface results in the change of the refractive index or of the local value of the dielectric constant, the change of reflectivity is observed. This phenomenon has been used as the mechanism for detection of gases (Fig. 9.18a) and of adsorbed biomolecules (Fig. 9.18b). The depth of penetration of the surface plasmon is comparable to that of the evanescent field, that is, 100–500 nm for the visible–near infrared range.

It is important to realize that for the plasmon resonance to occur the condition of two matching plasmons at the opposite interfaces of the thin metal must be met. In other words there must be a dielectric/metal interface at which an evanescent field is created. In the Kretschmann geometry that interface is created by having the metal coated on the glass prism. Likewise, the SPR condition can also be realized in a fiberoptic format with a thin metal layer deposited on a flattened single-mode optical
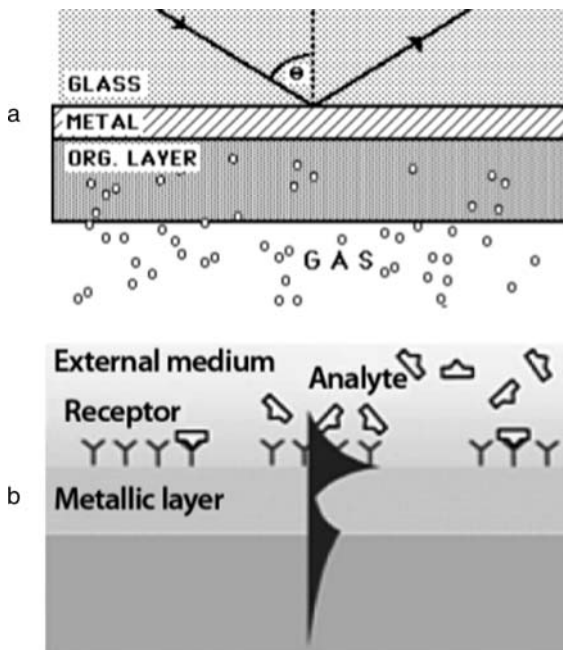
Fig. 9.18 Evanescent sensor for: (**a**) gases and (**b**) biomolecules (adapted from Lechuga et al., 2004)

fiber. In this case, the light propagates through the fiber, creates an evanescent field, and excites the plasmon at the metal film. Another option is plasmon originating in "surface grating." All these options have a common denominator: interactions of the complex components of the refractive index and the dielectric constant of the materials constituting the interfaces.

Related to the plasmon resonance physics is the "micromirror optical sensor for hydrogen" (Butler, 1991). Like gold and silver, palladium is a free-electron gas metal in which charge groupings such as phonons or plasmons are likely to occur. As we have seen already, palladium has a natural selectivity due to its sorption of monoatomic hydrogen. In that sensor, the reflectivity of the thin Pd film mirror mounted at the end of cladded optical fiber (Fig. 9.19) is modulated by absorption of hydrogen.

9.3.5 Photoluminescent Schottky Diodes

This sensing principle does not truly belong to the reflectance category. It is included here only because it does not fit better in any other place. Yet, its operation is interesting and important. It is a cross between the optical and electrochemical
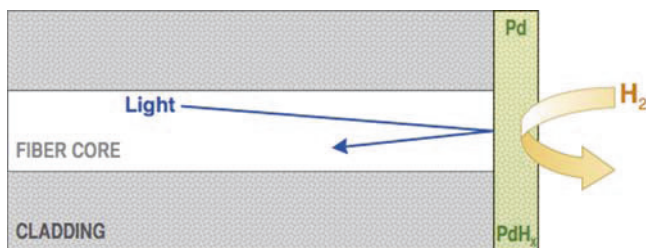
Fig. 9.19 Palladium micromirror sensor for hydrogen based on changes of reflectivity (adapted from Butler, 1991)

sensing principles. It has potentially wider applicability for sensing with junctions of organic semiconductors and metals.

Palladium has also been used in photoluminescent Schottky diodes and has been demonstrated as another type of hydrogen sensor. However, this sensing principle is more general and can be used for other gas-sensitive Schottky junctions. The space charge formed at such a junction originates from the partitioning of charge at the semiconductor/metal interface. It is chemically modulated through the changes of work function. Although this effect is primarily electrochemical, here it is detected optically.

When a semiconductor is illuminated with light of sufficient energy, electrons from the valence band are excited to the conduction band and the hole–electron pair is generated. The relaxation can take several pathways some of which are non-radiative. The radiative transition leads to photoluminescence (Fig. 9.20). This is a solid-state analogue of molecular fluorescence. The nonradiative recombination and radiative transitions are again the two competing processes.

A Schottky barrier junction is constructed by, for example, depositing Pd on n-CdS (Seker et al., 2000). Its simplified energy band diagram is shown in Fig. 9.21 together with a voltage source that applies a suitable bias across the diode.

In Chapter 8, we discussed the semiconductor–(insulator)–metal junction and the Schottky junction under conditions of a zero electric field, in the "flatband condition" (Potje-Kamloth, 2008). On the other hand, the Schottky junction (metal–semiconductor) used in photoluminescent studies is under electrical bias, just as the chemiresistor Schottky diodes. It means that the energy bands are bent and the region of nonzero electric field (space charge) is formed at the semiconductor side of the junction (Appendix C). Its presence is important for the effect to be used for chemical sensing. The hole–electron pair within this region is separated by the electric field, thus reducing the probability of the recombination process. Because the space charge is nonemissive, it is called a "dead layer" in the context of this discussion. In other words, the relative intensity of the emitted light is decreased as the thickness of the dead layer increases and vice versa. The semiempirical relationship describing this effect is formulated in terms of fluorescent quantum yields ζ_{air} and ζ_{H_2} for the illuminated junction in air and in hydrogen, respectively,
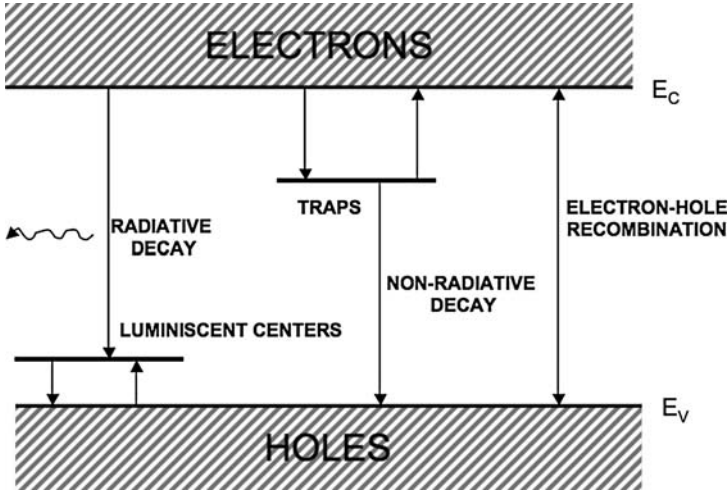
Fig. 9.20 Electron-hole transitions in wideband semiconductor leading to photoluminescence
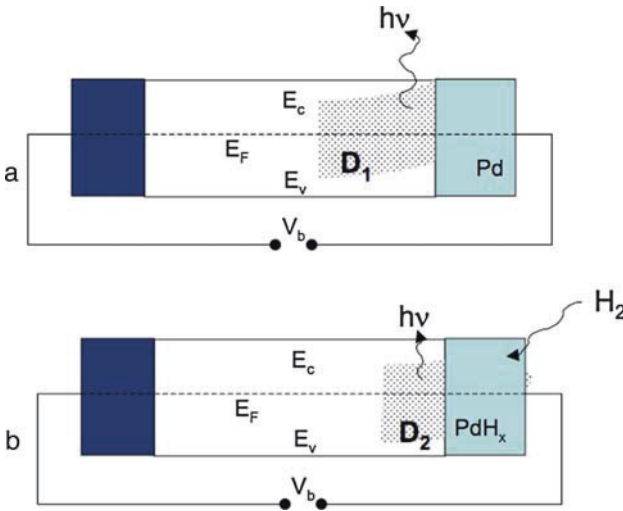
Fig. 9.21 Change of the "dead-layer" thickness $Dd = D_1 - D_2$ upon exposure of Pd to hydrogen in the photoluminescent Schottky junction

$$\frac{\zeta_{\text{air}}}{\zeta_{\text{H}_2}} = \exp - (\alpha + \beta)\Delta d \tag{9.40}$$

where α and β are the absorptivities of the semiconductor for the exciting and emitting radiation and Δd is the change in the dead-layer thickness. At a constant applied

bias, the thickness d of the dead layer region is given by (6.72). It is modulated by the change of work function of Pd (Section 8.1.5.2).

This concept can be extended to other semiconductors (Seker et al., 2000) in which the photoluminescent "dead layer" is modulated in a more subtle way. For example, n-GaAs derivatized with ferrocene responds to volatile oxidants (such as halogens) according to the above mechanism because the ferrocene molecule facilitates partial transfer of electron from the semiconductor to the adsorbed molecule. This separation of charge creates an additional electric field that affects the thickness of the dead layer at the surface and the intensity of photoemission from this layer (9.40). In principle any gas with electron donor–acceptor properties should yield a similar response. It is interesting to note the analogy of this mechanism with the one on which semiconducting oxide sensors (Section 8.3.2) and work function CHEMFETs (Section 8.2.2) are based. In semiconducting oxide sensors the modulation of the surface concentration of the adsorbed oxygen affects the thickness of the space-charge region and hence the conductivity.

9.4 Selectivity

It is safe to say that any conventional spectrophotometric and colorimetric analysis can be performed in an optical sensing format. That makes the optical sensors probably the most universal type of chemical sensors. Miniaturization of optical components and rapid advances in the development of new optical materials and hardware support this fast-growing area of chemical sensing.

In all chemical sensors, the selectivity originates partly in the physics of the transduction process and partly in the chemical interactions of the analyte with the selective layer. In different types of chemical sensors, these two selectivities always act together, but contribute in different proportions. Selectivity in optical sensors, more than in any other type of chemical sensor, relies on the physical selectivity of the transduction mechanism itself. Thus, electronic transitions in the UV–Vis part of the spectrum provide information about functional groups of the absorbing molecule whereas in the infrared (IR) region, the sensing information is obtained directly from the highly specific absorption at a single or multiple wavelength.

In chemically based selectivity, the sensing information is usually obtained from the "color reaction" between the analyte X and the indicator molecule I forming a colored complex $X * I$, in which either the indicator and/or the conjugate have characteristic absorption

$$X + I = X * I$$
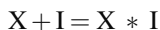
Many chromophore reagents are available for that purpose. Therefore, selectivity to a very large extent depends on the materials and optical components that allow us to deliver the light and manipulate it as needed in one or another optical sensing scheme.

9.4.1 Design of Optical Sensors

There are many different forms of optical sensors that differ in details of realization of the optical path. In this section, we highlight the essential elements of various implementations of optical sensors in which optical fibers and waveguides are used to pipe light around. From the spectroscopic point of view, there is mostly nothing fundamentally different in optical sensing that could not be done by performing the same experiment in a conventional arrangement, in a well-designed spectrophotometer. The convenience of remote sensing, electrical safety, the monolithic character of the optical probe, and all the other real and fictional benefits usually quoted in favor of optical sensors are often done at the expense of decreased accuracy and precision.

Inasmuch as one of the great advantages of optical sensors is their ability to perform remote sensing, waveguiding of light over distances ranging from centimeters to tens of meters is important. The parameter on which this ability hinges is the attenuation in the optical conduit (X). It is expressed in decibels (dB), which are dimensionless logarithmic units of the power ratio[1] defined as

$$X_{dB} = 10 \log_{10} \frac{X}{X_0} \qquad (9.41)$$

For optical waveguides, the frequency (or wavelength) at which the transmission occurs must be specified. In Table 9.2 the attenuations of the various materials used for construction of optical sensors and their usable ranges are given. The geometry of the waveguide also plays a role. As long as the bends in the fiber are not too sharp (the radius of the bend must be much greater than the wavelength) the direction of the propagation can be changed easily. Thus, light can be delivered to and from active sensing areas even if they are far from the source and not in the line of sight with the detector. This makes optical sensors particularly attractive for remote and hard-to-reach applications or where the use of electricity may be hazardous.

Let us now focus on the selective layer and how it is incorporated in optical sensors. Unless there is a specific reason for it, we use the term "optical fiber" in the general sense, meaning both optical fibers and flat optical waveguides. The selective layer can be placed on the fiber in several different ways depending on the type of interaction involved (Kuswandi et al., 2001). The general arrangement is shown in Fig. 9.22.

The reflecting surface can be a mirror or a membrane with a light-scattering surface. In any case, the sensor has the appearance of a monolithic probe (i.e., a "dip-stick" probe). Optical sensors based on absorption, fluorescence, phosphorescence, and luminescence can employ such a configuration. Various highly optimized fiberoptic probes for UV–Vis, NIR, and IR ranges are now commercially available, and their designs are shown in Fig. 9.23.

The second option is to make use of the continuous evanescent field and to locate the selective layer within this field. To that end, the protective coating and

---

[1] For the definition of attenuation of intensity, the multiplier in front of the log is 20 rather than 10.
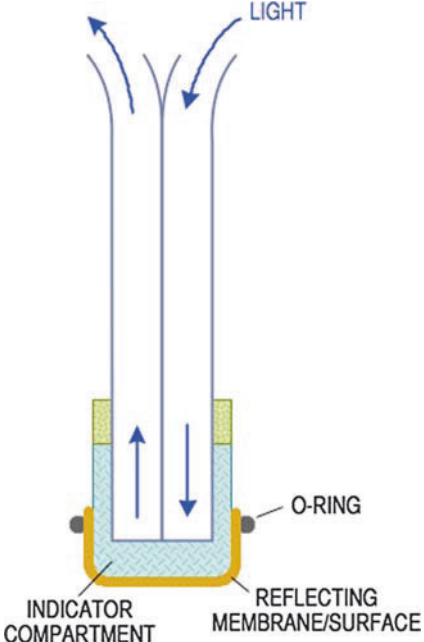
Fig. 9.22 General arrangement of fiberoptic sensor
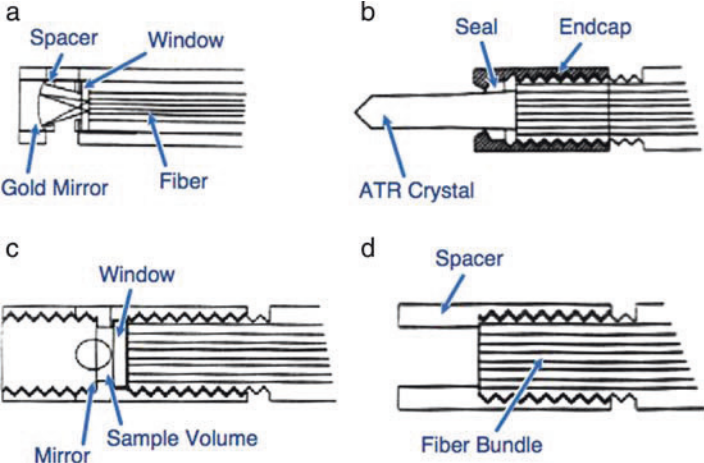
Fig. 9.23 Examples of infrared fiberoptic heads for sensing: (**a**) in gas; (**b**) by attenuated total reflectance; (**c**) in liquids; (**d**) by reflectance (adapted from Melling and Thomson, 2002)
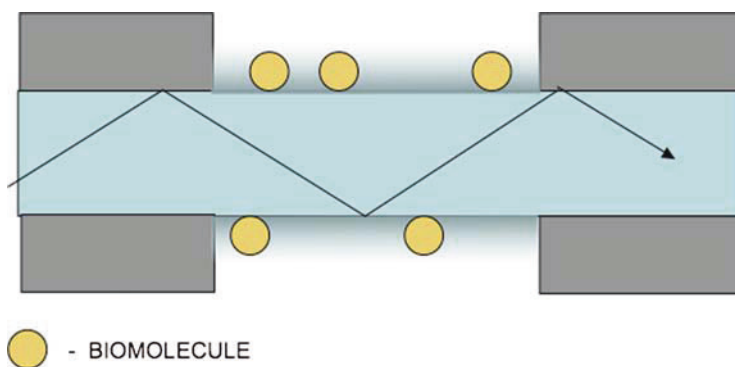
○ - BIOMOLECULE

Fig. 9.24 Schematic of evanescent field biosensor. The intensity of the light is attenuated by the interaction of the light with the molecules present at the surface of the waveguide

the cladding are replaced (Newby et al., 1984) with the selective layer (e.g., a biomolecule) in a short segment of the fiber (Fig. 9.24). The number of molecules of the absorber present within this evanescent field volume depends on the nature of the adsorption process itself. The "optical volume" in evanescent measurements is not constant, but depends on the wavelength, refractive indices of all materials involved, and on the angle of incidence (9.23). In these complicated interrelated processes, the concentration dependence of the signal is seldom calculated, but is rather established as an empirical calibration curve for the given standard set of experimental conditions.

Hollow WaveGuides (HWGs) represent a qualitatively different approach to optical sensing. Here, the HWG serves as the light conduit, and the interaction with the chemically selective indicator takes place either inside the hollow core of the HWG or on the outer surface of the waveguide in the evanescent field. The optical waveguide sensor for NH_3, in which the indicator oxazine perchlorate was used as solid film on the outer surface of a glass capillary tube (Fig. 9.25), is an early example of such an HWG/evanescent wave sensor (Giuliani et al., 1983). Particularly advantageous use of HWG is in the mid-IR range, because transparent materials are still rather limited and/or expensive (Mizaikoff, 2003; Charlton et al., 2005).

In this version, the radiation again passes through the hollow core of the capillary but interacts with the sample inside the capillary volume. Therefore, it is a modern analogue of a multiple-pass "gas IR cell." Because light is multiply reflected from the inner wall of the HWG, the actual optical path is much longer than the physical length of the HWG. The obvious question with such "gas cell" HWG sensing is the accessibility of the interior for the sample, as it affects the response time. An unaided diffusion would take too long. Therefore, some form of convective transport (a pump) is always needed. This means that the inlet to the HWG can be used for point sampling ("sniffing") of the local concentration.

Finally, there is another possibility of arranging interaction between light and the analyte. It is called open optical path sensing. Here, no physical waveguide is used. The beam of light is emitted from a coherent, narrow bandwidth source and
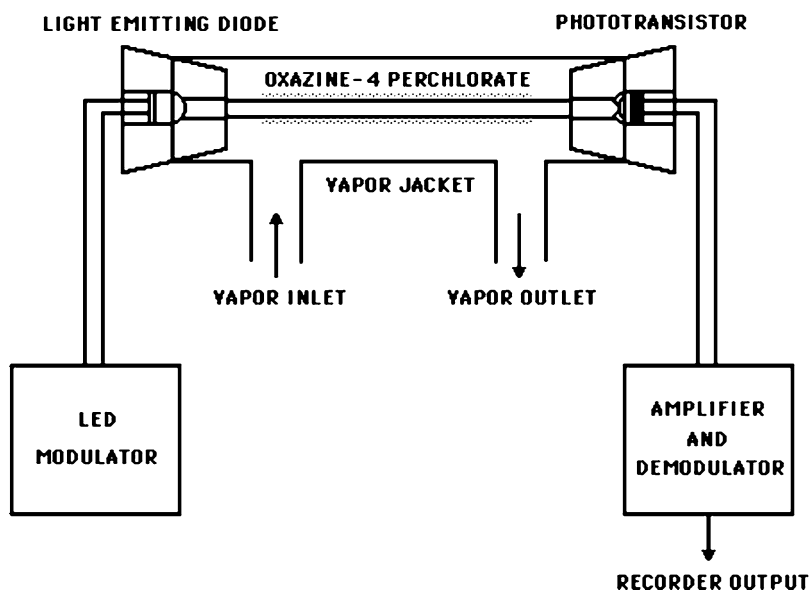
Fig. 9.25 Hollow waveguide for ammonia, based on capillary coated with oxazine perchlorate (adapted from Giuliani et al., 1983)

the attenuation of its intensity is measured by the detector, mounted at some fixed distance away. This form of sensing has been used for detection of "fugitive vapors" from contaminated environmental sites at which a large area of land needs to be continuously monitored. The Quantum Cascade Lasers (QCL), with their narrow emission lines in the mid-IR range and high power density, are ideal light sources for that kind of application (Faist et al., 1994).

A direct analogue of the Severinghaus electrode (Chapter 6, Section 6.2.2) or its conductometric version (Chapter 8) can be made by placing an internal optical sensing element behind a membrane. Sensors of this type have been made and tested for response to NH_3 and the extension to other acidobasic gases (Arnold and Ostler, 1986). The internal solution contains a suitable acidobasic indicator. The dynamic range of this sensor can be adjusted by the concentration of the indicator, which acts as its own buffer. Similar to the Severinghaus, an electrode membrane sensor of this type can be used either in liquid or in gas samples with interferences coming only from other acidobasic gases. Because the composition of the solution in the inner compartment is constant, the activity coefficient ratio and the activity of water are constant. Therefore there are no problems due to their variation. It should be noted that the inner solution can be any solvent, including one that is totally nonconducting.

A general challenge of liquid optical sensing comes from the effect of the medium. Obviously, if the medium is strongly absorbing in the same optical range as the analyte, the signal-to-noise ratio is seriously affected. This is a general problem facing application of mid-IR-range sensors in water-containing media.

Broad absorption of IR radiation in water (at 3,300, 2,100, 1,640, and $750\,cm^{-1}$) makes sensing of organic species difficult. Under normal circumstances, it is limited to the high concentration range. This problem can be overcome by coating the ATR element with a hydrophobic material that extracts the lipophilic constituents from water and at the same time drastically limits the presence of water at the interface. It is an idea borrowed from the "Solid Phase Extraction" (SPE) sampling used in gas chromatography and mass spectroscopy. In order to be effective, such an "enrichment membrane" (Mizaikoff, 2003, 1995) has to be thicker than several multiples of the penetration depth (9.19). Its chemical nature and the exact thickness are matters of compromise between the suppression of the water background and the response time (Fig. 9.26).

Thus, a polyisobutylene coating on silver halide optical fiber was used for environmental analysis of chlorinated hydrocarbons in water in an instrument designed for operation in seawater 500 m under the surface (Fig. 9.27). The enrichment membrane concept can be used for FTIR-ATR analysis of liquid and gaseous samples alike.
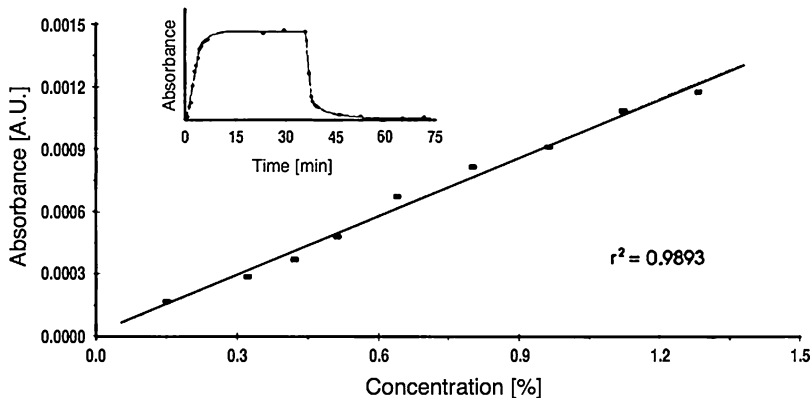
Fig. 9.26 Enrichment PIB membrane (adapted from Miizaikoff, 1995)
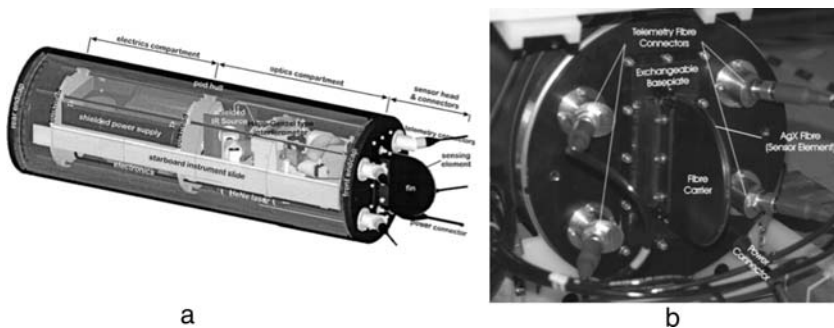
Fig. 9.27 Fully contained ATR/FTIR instrument rated for operation at the depth of 500 m in the ocean. The ATR sensing element is AgCl fiber coated with polyethylene/polypropylene copolymer (reprinted from Kraft et al., 2003)

9.4.2 Optical Arrays

A multitude of available optical reagents for both liquid and gaseous samples, combined with different modes of optical sensing (wavelength, fluorescence, etc.) invites the realization of optical arrays. These can be implemented in a deterministic format; that is, each sensing element has a predetermined address in the array (Zhang and Suslick, 2005). A combination of reagents based on charge-transfer complexes, hydrogen bonding, and Bronsted acidity have been assembled in a 6×6 array of dots (Fig. 9.28; Suslick, 2004) and exposed to various amines in buffer solution. The color pattern that developed has been evaluated by pattern recognition techniques (Chapter 10).

A similar strategy has been adopted for multivariate detection of vapors with a set of metalloporhyrins with different central metal atom (Rakow and Suslick, 2000). They act as solvatochromic indicators; that is, their spectrum changes with the vapor molecules solvating the indicator.

If the responses are sufficiently orthogonal, the information content of the array can be enhanced by increasing the number of array elements. Also, from the point of view of statistics, a certain amount of redundancy can increase the precision. A large number of array elements immediately raise the issue of addressability. That and the method of array fabrication have been elegantly solved by D. Walt and his group (Michael et al., 1998). In that case, the flat end of the imaging fiber bundle, consisting of thousands ($\sim$20,600) of individually clad optical fibers is chemically etched, forming an array of hexagonally packed "microwells." Several sets of beads chemically modified with suitable optical reagents are prepared. The beads are monodispersed (i.e., the same diameter) and can be made of a variety of materials, typically glass or polystyrene. Any number of such chemically modified beads are then mixed and exposed to the etched end of the fiber. The electrostatic attraction causes the beads to embed themselves randomly in the microwells (Fig. 9.29).

In the next step, the bead-modified fiber bundle is sequentially exposed to analyte-1, analyte-2, and so on. At each exposure, the visual image of the array is recorded and stored. The result of this encoding step is a response map, which is
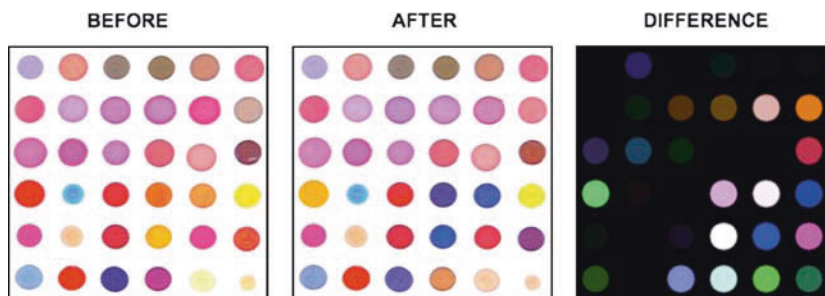
Fig. 9.28 Color changes in 6×6 array of solvatochromic, pH-sensitive and shape-recognition dyes when exposed to ammonia (adapted from Suslick, 2004)
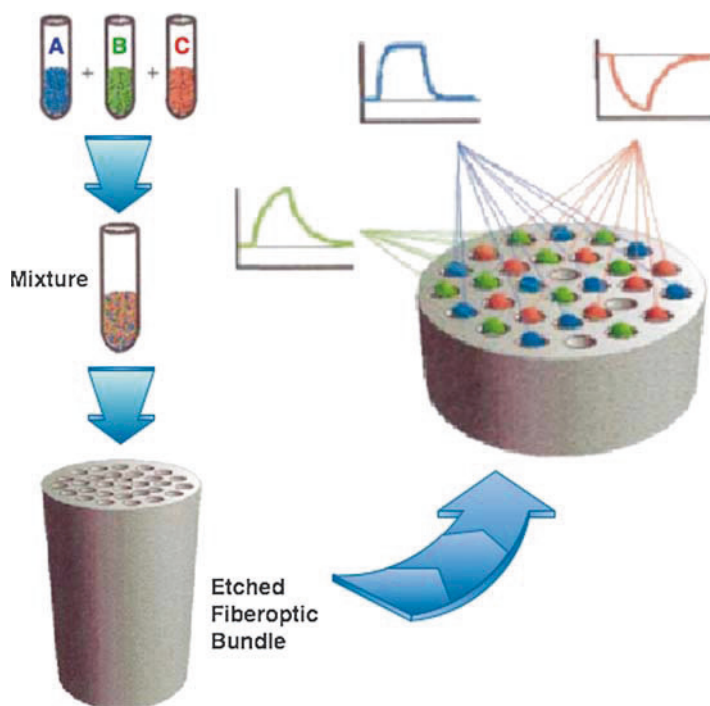
Fig. 9.29 Randomly ordered, addressable, high-density optical sensing array (see text for description of steps; adapted from Michael et al., 1999)
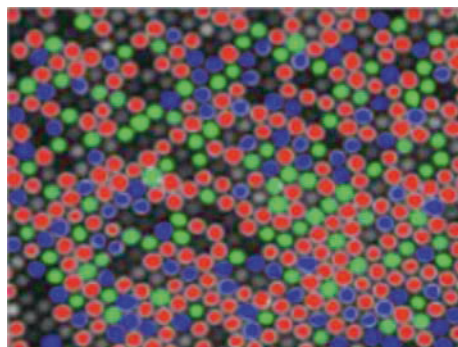
Fig. 9.30 Color map obtained from the array in Fig. 9.27 (adapted from Michael et al., 1999)

then used to decode the response to the analyte (Fig. 9.30). This ingenious approach has been used in combination with multivariate analysis (Chapter 10) for fluorescent detection of vapors (Albert and Walt, 2000) and for affinity-based detection, based on aptamers as reagents (Lee and Walt, 2000).

9.4.3 Optical Ion Sensors

In addition to the transparency of the media, there is also an issue of the inter-
action of the analyte itself with the medium and how it affects the sensing of
ionic species, in so-called optrodes. Most ions in aqueous solution do not have a
characteristic spectrum that would allow their direct measurement. Therefore, their
sensing is performed via the "color reaction" with some specific reagent immobi-
lized in the selective layer. Various complexometric reagents and color indicators
are used for that purpose. These interactions are described in terms of equilibria
which, from the thermodynamic point of view, must be written in activities and not
in concentrations (Appendix A). This is required by the nature of the Coulombic
interactions of the ions with the medium (Table 2.1). The electromagnetic radiation
interacts with matter on an atomic and molecular level. In other words, light counts
molecules and therefore provides information about concentration. Interactions of
the molecules among themselves and with their environment, which accounts for
their activity, usually appear only as secondary effects (e.g., shift of the absorption
and/or emission peak, line broadening, change of fluorescence decay time, etc.).
Furthermore, there are no general quantitative rules that connect these secondary
effects with the activity coefficient of the molecule in the given environment and the
molar absorptivity of the ion. Thus, there is a fundamental discrepancy between the
transduction mechanism and the origin of selectivity provided by the color reaction
which is inherent in optical sensors.

Optical sensors for ions use indicators, which exist in two different colors,
depending on whether the analyte is bound to them. The use of colored indicators is
one of the oldest principles of analytical chemistry, used extensively both in direct
analytical spectroscopy and in so-called visual titrations. In their sensing applica-
tion, the indicator is confined to the surface of the optical sensor or immobilized in
the selective layer. In that sense, the oldest and most widespread optical sensor is
a pH indicator paper, the litmus paper, which is commonly used for the rapid and
convenient semiquantitative estimate of pH of solutions or for endpoint detection in
acidobasic titrations. Its hi-tech counterpart is a pH optrode (the name of which is
intentionally reminiscent of the pH electrode), which essentially does the same thing
(Wolfbeis, 2004). The operation principles and limitations of ion optical sensors are
common for all ions.

The primary chemical interaction governing the operation of the pH optrode is
the acid-base equilibrium of the indicator **HA**.

$$HA + H_2O = H^+ + A^-$$
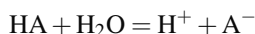
Its dissociation constant K_a is defined by the equilibrium equation, which has been
arranged such that it is usable for optical sensing:

$$K_a = \frac{a_{H^+} a_{A^-}}{a_{HA} a_{H_2O}} = a_{H^+} \frac{C_{A^-}}{C_{HA}} \frac{f_{A^-}}{f_{HA} f_{H_2O}} \tag{9.42}$$

In logarithmic form, it is known as the Henderson–Hasselbalch equation.

$$\text{pH} = \text{p}K_{\text{a}} + \log \frac{C_{A^-}}{C_{\text{HA}}} + \log \frac{f_{A^-}}{f_{\text{HA}}} - \log a_{\text{H}_2\text{O}} \qquad (9.43)$$

The second term shows the log of the ratio of concentrations of the protonated **HA** and dissociated **A$^-$** form of the indicator, which are measured optically. If we assume that a hundredfold change of concentration of the absorbing molecule can be conveniently measured (i.e., the dynamic range belonging to one K_{a} equals ± 1 decade), which means that one indicator can cover two units of pH, by selecting a series of indicators with suitably spaced values of their dissociation constants, a pH optrode with broad dynamic range can be obtained.

The third and fourth terms in (9.43) are usually ignored. This can be accepted for very dilute aqueous solutions where the activity coefficients tend to unity, their ratio converges even faster with dilution, and water is in a larger molar excess. Unfortunately, conditions in most real measuring situations are not such that these terms can be ignored with confidence. The variability of these two terms then introduces error into the optical determination of pH or other ions.

The effect of the environment on the dissociation equilibrium of the indicator used in optical ion sensors is through the ratio of the activity coefficients of its neutral and charged forms (the third term in (9.43)). These effects include the change of the ionic strength, specific interactions due to, for example, proteins, adsorption effects, and the like. Attempts to account for some of these interactions on the basis of the classical Debye–Huckel theory have been made and an optrode for determination of ionic strength has been constructed (Wolfbeis and Offenbacher, 1986). This approach works reasonably well for uni-univalent electrolytes up to ionic strength $I < 0.1 \,\text{M}$. For higher ionic strengths and/or for multivalent ions, the specific interactions must be taken into account (Janata, 1987). These corrections involve the ionic sizes of all ions present in the medium and the type of the indicator used.

The effect of proteins on the dissociation constant of indicators (often called protein error) is well known. It is defined as the difference between the colorimetrically determined values of pH in the presence and in the absence of the proteins in solution, whose pH has been adjusted (usually electrochemically) to its original value. Depending on the type of proteins, their concentration, and on the type of the indicator, it can be as high as 0.8 units of pH. Again, it has its origin in the specific interactions between the indicator molecule and the proteins.

There have been several attempts to overcome these problems. It follows from the above discussion that the relative contribution of the third and fourth terms in (9.43) to the observed value of pH depends on both specific and nonspecific interactions. That contribution, in turn, depends on the type of the indicator used, where and how it is immobilized, and on the type and concentrations of all other ions in the sample. Thus, in one approach (Opitz and Lubbers, 1983) two different types of indicators have been used and measured at two different wavelengths. Because of both specific and nonspecific interaction, terms for these two indicators are very different. Their dependence on the ionic strength is also sufficiently different to yield two sets of equations that are necessary to solve for both pH and ionic strength.

The problem is that those interaction terms are valid only for one specific type of ion (both charge and chemical type) and if the general composition of the sample changes, such corrections are no longer valid.

When the indicator is located at the surface of the waveguide, it responds to the local pH. In order to mitigate the problem of changing local ionic strength, it has been suggested to place the indicator in a highly charged environment, by surrounding it with surface immobilized quarternary ammonium groups (Wolff-beis and Offenbacher 1986). This, however, creates another kind of problem. The color indicator which is confined to the surface of the optical fiber must equilibrate between the bulk of the sample and this surface region. Thus, (9.43) must be written in surface activities. In other words, the dissociation equilibrium is based on the conditions as they exist at the sensor interphase (surface and the adjacent region). It means that the above discussion, including the variations in ionic strength, the protein effect, and so on, must be based on the conditions at the interphase. Most proteins will be present at the interphase in concentrations that far exceed their bulk value. Also, the activity of water will be determined by the hydrophobicity of the interphase; ions will specifically adsorb, and so on.

It is a well-known fact that upon covalent immobilization at the surface, the dissociation constant of the acid-base indicator changes by as much as 3 pK units. This shift clearly illustrates the dramatic effect that the interphase has on the ionization equilibria. It is perhaps the most serious problem with optical sensors: that the surface concentration of any species is related to its corresponding bulk activity value through an adsorption isotherm which, with the exception of Henry's law, is a highly nonlinear and variable relationship. It is also known (Davies and Rideal, 1963) that the surface pH is different from the bulk value due to the electrostatic repulsion.

$$pH_{surf} = pH_{bulk} + \frac{Ne\phi_s}{2.3\Re T} \qquad (9.44)$$

The surface potential ϕ_s depends on the concentration profiles of all ionic species present at the interphase, including ionic strength affecting the diffuse layer, specifically adsorbed ions, and surface ionizable groups. Thus, the surface itself possesses acid–base properties that are reflected in the response of the immobilized indicator dye but are not uniquely linked to the pH of the bulk of the solution.

So far, we have ignored the last term in (9.43), the logarithm of activity of water. This is acceptable in aqueous solutions, however, in mixed solvents or in concentrated electrolyte solutions, this term is dominating. This phenomenon is common, to a varying degree, for all water–organic solvent mixtures, and it is due to the fact that the activity of free uncoordinated water decreases as the concentration of the organic solvent increases. Because in solvation equilibria, all species are interdependent, it is not surprising that the activities of all species are affected in a most profound way by the presence of organic solvents. The degree to which individual ions, including the indicators, are affected by the medium depends on their structure, which gives rise to a multiplicity of acidity functions obtained for structurally different types of indicators (Boyd, 1969).

The activity of water plays a dominating role in optical ion sensors in which the indicator dye is incorporated in a hydrophobic layer. Despite its "hydrophobic" label, such layers contain a finite, sometimes very significant concentration of water. In that respect, they behave as mixed organic–aqueous media with all the implications that affect the aforementioned acidity functions. In other words, the optical signal coming from such a selective layer is affected more by the degree of hydration than by the changes of ion activity in the solution (Janata, 1992).

It has been proposed to base ion optrodes on the Stark effect that the electrostatic field at the membrane/solution interface creates in field-sensitive dyes (Wolfbeis and Schaffar, 1987). This effect relates the frequency shift Δv of the permanent dipole μ of the dye to the local electric field, according to the equation

$$\mu = \frac{\Delta v E}{hc} \tag{9.45}$$

The electric field is generated at the interface by the selective partitioning of the ion from solution to the membrane. Therefore, the process is fundamentally electrochemical, based on the Nernst potential, but only the readout is optical. In principle, this could again lead to sensing of ion activity without a reference electrode! The problem with this scheme is that the electrostatic potential gradient (i.e., the field) at the interface is affected not only by the field originating from the Nernst potential, but by the adsorption of charged and dipolar species, and by the gradients of the dielectric constant at both sides of the interface. Furthermore, the electric field is a vector (Fig. 9.30). This means that in order to have a defined polarization effect, the dye would have to be uniquely oriented. This has been attempted by immobilizing such dyes in oriented Langmuir–Blodgett multilayers (Schaffar and Wolfbeis, 1988).

Another factor to consider is the distribution of the dye in the selective layer. For practical reasons, the selective layer would be solvent-cast on the optical fiber. It would contain the appropriate dye or dyes that would generate the desired optical response upon complexation with the target ion. Ionopohores with suitable chromophores attached to them, so-called chromoionophores, have been synthesized for this purpose (Wang et al., 1990). Membranes with a thickness of a few micrometers to tens of micrometers have to be used in order to obtain a measurable signal. As the analyte ion diffuses into such a layer, it progressively complexes the chromoionophore. This results in unusually long response times, on the order of hundreds of seconds.

The above discussion centered around pH. It served the purpose of highlighting general problems that affect all optical sensors for ions. In spite of these problems, optical sensing of ionic species has some attractions. The main one is the fact that determination can be done with a monopolar probe (i.e., without a reference electrode required for an electrochemical measurement). The price to be paid for such an advantage is a significantly lower accuracy of such measurements in all but very dilute aqueous solutions and much greater vulnerability to nonspecific interferences. For this reason, the direct optical sensing of ions is only semiquantitative, at best. The experimentally determined pH errors due to the liquid junction

of ISE are always smaller than those caused by the effect of ionic strength, organic solvent, or proteins on optical indicators. This is the main reason why conventional measurement of pH and other ions by electrochemical methods is preferable to the optical one.

9.4.4 Fluorescence Gas Sensors

The above problems have their origin in the strong interaction that ions have with their environment. However, they do not appear in optical sensors for electrically neutral species, which are generally free of the above problems. The number of possible combinations of different chemistries and different optical arrangements is obviously very large. No attempt is made here to present an exhaustive review and the reader is referred to large reviews dedicated to this subject (e.g., McDonagh et al., 2008).

Sensors based on fluorescence are quite robust because the wavelength and the orthogonal detection geometry of the incident and emitted radiation results in a high signal-to-noise ratio. The sensors described here utilize quenching of fluorescence. Thus, the analyte is the quencher Q and the indicator is a fluorescing dye F, which when excited to F^*, emits fluorescence with a characteristic decay time.

$$F + h\nu_1 = F^*$$
$$F^* \overset{\tau}{\Rightarrow} F + h\nu_2$$
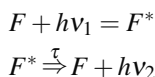
The competing reaction is reversible binding of the quencher by the excited fluor F^*.

$$F^* + Q \overset{K_{SV}}{\Longleftrightarrow} FQ$$
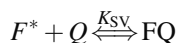
The quenching equilibrium is described by the fluorescence quenching constant.

$$K_{SV} = \frac{C_{FQ}}{C_{F^*}C_Q} \tag{9.46}$$

The total fluor concentration C_F is

$$C_F = C_{F^*} + C_{FQ} \tag{9.47}$$

Only the "free" excited fluor emits light. Equations (9.10), (9.46), and (9.47) then yield the Stern–Volmer equation,

$$\frac{I_0}{I} = 1 + K_{SV}C_Q \tag{9.48}$$

in which the I_0 is the fluorescent intensity at zero concentration of the quencher and I is the intensity of "free" fluor F*. A plot of I_0/I against the concentration of

the quencher is a straight-line calibration plot, with Stern–Volmer constant K_{SV} and intercept $+1$.

By far the most important quencher is molecular oxygen (Demas et al., 1999). The oxygen sensor based on fluorescence quenching has performance rivaling the Clark amperometric electrodes. Many fluors have been synthesized with fluorescent lifetime in tens of nanoseconds. Fluors with lifetimes shorter than 10 ns do not have high enough fluorescence yield ζ and are not generally suitable for sensing purposes.

If several quenchers are involved (9.48) expands to

$$\frac{I_0 - I}{I} = K_1 C_{Q1} + K_2 C_{Q2} \dots \tag{9.49}$$

Using at least as many sensors with different Ks as there are quenchers, one can obtain enough equations to determine the mixture. This approach has been used for the determination of mixtures of halothane (2 chloro-2-bromo-1,1,1-trifluoroethane) and oxygen (Wolfbeis et al., 1985). Simultaneous monitoring of these two gases is important in management of general anesthesia. Two sensors **A** and **B** with different values of the quenching constant K_A and K_B have been constructed and their relative fluorescences (9.49) measured in pure quencher gas are α and β, respectively. By simple algebra, the concentration of halothane C_H and oxygen C_O can be expressed as

$$C_H = \frac{\alpha K_B^O - \beta K_A^O}{K_A^H K_B^O - K_A^O K_B^H} \tag{9.50}$$

$$C_0 = \frac{\beta K_A^H - \alpha K_B^H}{K_A^H K_B^O - K_A^O K_B^H} \tag{9.51}$$

Halothane can be effectively blocked by a 12 μm thick Teflon membrane (Fig. 9.31).
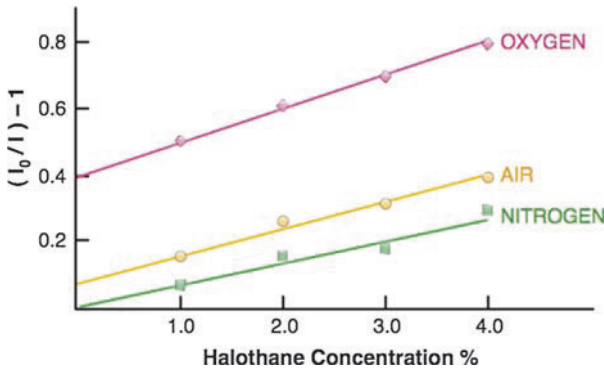
Fig. 9.31 Response of fluorescent gas sensor with Teflon membrane, according to (5.49) (adapted from Wolfbeis et al., 1985)

When such a membrane is used to shield the sensor **B**, the above equations simplify to

$$C_H = \frac{\alpha K_B^O - \beta K_A^O}{K_A^H K_B^O} \tag{9.52}$$

and

$$C_0 = \frac{\beta}{K_A^H} \tag{9.53}$$

If the Stern–Volmer constants for oxygen for both sensors are the same, (9.52) simplifies to

$$C_H = \frac{\alpha - \beta}{K_A^H} \tag{9.54}$$

Several different fluors have been tested for this application and the best results have been obtained for decacylene ($\lambda_{ex} = 385\,nm$; $\lambda_f = 510\,nm$; $K_{SV} = 0.62$) embedded in silicone rubber film. Fluoranthene which has a higher value of K_{SV} requires UV excitation ($\lambda_{ex} = 360\,nm$) and is therefore less convenient to use.

The response time is between 15 and 20 s for both gases, and the precision is better than that obtained with the Clark electrode. A sterilizable oxygen probe based on this principle with performance characteristics equal to or better than the amperometric oxygen electrode has been constructed (Kroneis and Marsoner, 1983).

9.4.5 Optical Biosensors

The combination of selectivity of biochemical origin with the high intrinsic selectivity of optical spectroscopy makes the optical sensors potentially the most selective of all types of chemical sensors. It is not surprising, because most bioassays use optical detection. Moreover, the coupling requirements between the primary interactions in the selective layer and the transducer part of the sensor are relatively simple. The molecule that provides the signal (absorber) merely has to be present in the optical path in order to be counted.

In the actual implementation of optical biosensors, any optical principle discussed in Sections 9.2 and 9.3, or their combination, can be used. It is sometimes difficult to distinguish between spectrophotometric determination done in a conventional way (in the cuvette) and an identical assay in which the light is guided around by an optical fiber or a waveguide. In principle, direct (affinity) binding of biomolecules, such as DNA hybridizations, antibody/antigen immunochemical reactions, and receptor/agonist binding can be detected by one or another optical means (Borisov and Wolfbeis, 2008).

The detection and quantification of the presence of biomolecules at the surface is based on specific interactions taking place in the evanescent field, generated by the total internal reflectance or by the surface plasmon resonance. The latter is the key transduction principle in the optical bioanalysis and biosensing area (Narayanaswamy and Wolfbeis, 2004). Launched in the early 1980s in Sweden,

SPR instruments of Biacore AB (in 1996) became the industry standard. Although the primary transduction mechanism, modulation of the surface refractive index, is nonselective, the selectivity and versatility of these surface-based systems rely on the immobilized recognition sites. Specific "biosensor chips" are available to perform highly specific bioassays. Although one should never argue against success, that label, more than anything else, has helped to blur the line between the definition of a true "sensor" and a "bioassay." Because the binding in the main recognition event is so strong (see Chapter 2), these "biosensor chips" are not reversible chemical sensors. In fact, it can be said that there are no direct-binding biosensors, contrary to the hundreds of papers making that claim in the title. Such biosensors would be possible only if the value of the binding constant of the biorecognition event could be lowered.

With the above "binding" caveat, we make a brief general statement about direct-binding, affinity biosensors. A particular advantage of evanescent wave sensing is the measurement of fluorescence. This can be done in two ways. In both cases, the primary light is delivered to the selective layer zone through the fiber and the fluorescence is measured either in the perpendicular direction to the fiber (e.g., with the aid of another fiber), or it is coupled back to the same fiber. The longer wavelength fluorescence is separated from the exciting radiation with the help of some form of monochromator. The technique of Total Internal Reflectance Fluorescence (TIRF) has been applied to the study of adsorption of proteins using attached fluorescent labels (Sapsford and Ligler, in Wolfbeis, 2004) or directly by using the intrinsic fluorescence of the absorbing molecules (Hlady et al., 1985). Both alternatives offer the advantage of rejecting the background fluorescence from the bulk of the sample; the internal coupling of the fluorescent light has the advantage of signal enhancement due to the higher reported collection efficiency when the fluorescence is excited within the evanescent field. As a transduction principle, the evanescent field and SPR approach has great potential for the development of true biosensors. Nevertheless, the issue of sensing reversibility remains the dominating problem.

Enzymatic reactions coupled to optical detection of the product of the enzymatic reaction have been developed and successfully used as "reversible" optical biosensors. By definition, these are again steady-state sensors in which the information about the concentration of the analyte is derived from the measurement of the steady-state value of a product or a substrate involved in highly selective enzymatic reaction. Unlike the amperometric counterpart, the sensor itself does not consume or produce any of the species involved in the enzymatic reaction; it is a zero-flux boundary sensor. In other words, it operates as, and suffers from, the same problems as the potentiometric enzyme sensor (Section 6.2.1) or the enzyme thermistor (Section 3.1). It is governed by the same diffusion–reaction mechanism (Chapter 2) and suffers from similar limitations.

In this section, only the important differences resulting from the optical detection are highlighted. The optimum conditions for the operation of optical enzyme sensors, particularly the thickness of the enzyme layer, could be found again by solving the set of diffusion–reaction equations for a given geometry. There is, however, an
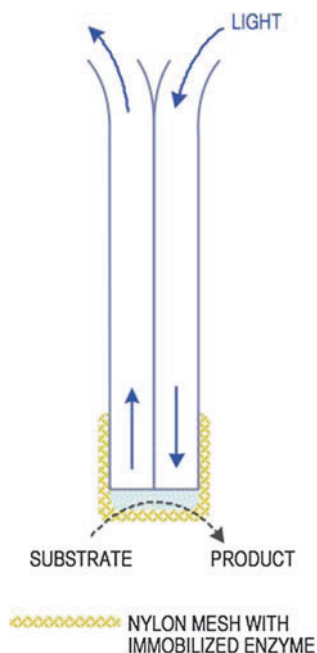
Fig. 9.32 General scheme of enzyme-based optical sensor

important difference: whereas in the potentiometric sensor it is only the concentration of the indicator ion at the interface of the electrode that generates the signal, in the case of the optical sensor, it can be the surface and/or the bulk of the entire enzyme layer. Obviously, because the number of the absorber molecules in the bulk is higher than at the surface, the optical enzyme sensor can in principle achieve a lower detection limit and a higher sensitivity than the corresponding potentiometric enzyme sensor.

An example of an optical enzyme sensor (Arnold, 1985) in a bifurcated optical fiber is shown in Fig. 9.32. The bifurcated fiber delivers and collects light to and from the site of the enzymatic reaction. The enzyme, alkaline phosphatase (AP), catalyzes hydrolysis of p-nitrophenyl phosphate to p-nitrophenoxide ion which is being detected ($\lambda = 404\,\text{nm}$).

$$\text{p-nitrophenyl phosphate} + H_2O \Rightarrow \text{p-nitrophenoxide} + \text{phosphate}$$
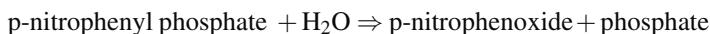
The enzyme is immobilized on a nylon mesh, which also acts as a diffuse reflector for the light. The dynamic range of this sensor is between 10^{-5} and $10^{-3}\,\text{M}$. Although the primary process that determines the steady-state concentration of the p-nitrophenoxide ion is the diffusion–reaction mechanism (which is governed by concentrations of all participating species), the detection of its concentration is again subject to the limitations of optical sensing of ionic species (Section 9.4.1). There are many similar optical enzyme biosensor schemes that utilize detection of

the colored product of the enzymatic reaction (Kuswandi et al., 2001). In addition to the detection of substrates, sensors for inhibitors and activators have also been described. Because they require the substrate for their operation, they have a finite lifetime and again fall into the domain of bioassays.

A sensor with glucose oxidase immobilized on top of a fluorescence quenching oxygen sensor shows excellent performance characteristics (Rosenzweig and Kopelman, 1996). It contains tris(1,10-phenanthroline) ruthenium(II) chloride immobilized in a heavily cross-linked acrylamide gel. The active end of the sensor has a diameter of only $0.5\,\mu m$, and the selective double-layer is formed by consecutive dip-coating/cross-linking of the fluor-containing layer and glucose oxidase. The small diameter of the probe tip apparently accounts for the exceptionally short ($<1\,s$) response time to a step change of glucose concentration in solution. It is possible that the radial diffusion geometry discussed in the context of the amperometric microelectrode (see Section 7.2) accounts for this outstanding performance.

Another interesting feature of this design is the chopped excitation light to $\sim 500\,ms$, which minimizes the photobleaching of the dye. It is difficult to pronounce judgment on the merit of the optical versus amperometric glucose sensing, because the choice would be always dictated by the circumstances of the real application. Nevertheless, it can be concluded that the two approaches are at least comparable.

Food for Thought #9

Activities

Optical sensors for ions have been called "opt(r)odes" in a rather suggestive analogy with "electrodes". If such an analogy were true, the problem of the reference electrode in electrochemical determination of ion activities will have been solved.

1. Explain why this cannot be so.
2. What defines the thermodynamic limits within which the "opt(r)odes" can be used to measure concentration of ions reliably?
3. Which optical sensor would be more affected by change of the concentration of electrolyte, glucose optical sensor, or calcium ion optical sensor? Explain why.

Evanescent Field

A guided light penetrates to the optically rarer phase (e.g., cladding), forming an evanescent field.

4. Explain the connection between the attenuation loss of the optical fiber used for transmission of information and the sensing principles utilizing an evanescent field.
5. Are there some "chemical effects" that would interfere with information transmission?
6. Could you use a graduated refractive index fiber for evanescent field sensing?

Optrodes in Dilute Solutions

In the optrode for calcium ion, a so-called "optical ionophore" is immobilized in a hydrophobic polymer matrix. It is a chromophore that changes its absorbance upon complexation with Ca^{2+}.

7. It has been suggested that such a sensor can be used for the accurate measurement of Ca^{2+} activity in dilute sample solutions because, in dilute solutions, activity coefficients tend to unity. Is that statement correct or not?
8. Does the signal in an opt(r)ode for calcium ion depend on the thickness of the membrane containing the calcium chromophore? How does it compare with the thickness dependence of the signal in a calcium ion-selective electrode containing calcium ionophore?
9. Is there a fundamental difference in the nature of the time response of these two types of sensors?

Symbols

A	Absorbance
C	Concentration
c	Speed of light $(2.9979 \times 10^8\,\mathrm{m\,s^{-1}})$
d_p	Depth of penetration
E	Electric field intensity
ΔE_λ	Quantum of energy
h	Planck's constant $(6.625 \times 10^{-27}\mathrm{erg\ s})$
I	Intensity
K_A	Quenching constant for species A
K_{SV}	Stern–Volmer constant
L	Pathlength
M	Number of modes
N	Number of photons/unit area
$N.A.$	Numerical aperture
n	Refractive index
P	Probability of absorption

r	Radius
S	Cross-sectional area
S_r	Scattering constants
V	"V number"
v	Velocity
ε_m	Dielectric constant of a metal
ε_λ	Molar absorptivity
εC	Absorption constant
γ	Collection factor
λ	Wavelength
Θ	Angle
ν	Frequency of oscillation
$\overline{\nu}$	Wave number
ζ	Quantum efficiency or quantum yield
ϑ	An angle

References

Albert, K.J. and Walt, D.R. (2000) *Anal. Chem*. 72, 1947.

Arnold, M.A. (1985) *Anal. Chem*. 57, 565.

Arnold, M.A. and Ostler, T.J. (1986) *Anal. Chem.* 58, 1137.

Beveridge, H.A. and Carlyon, R.P. (1996) *Hear. Res*. 99, 110–118.

Borisov, S.M. and Wolfbeis, O.S. (2008) *Chem. Rev*. 108, 423–461

Boyd, R.H. (1969) Interactions in dipolar aprotic solvents. In: Coetzee, J.F. and Ritchie, C.D. (Eds.) *Solute-Solvent Interactions*, Vol. 1. Marcel Dekker.

Butler, M.A. (1991) *J. Electrochem. Soc*. 138, L46.

Charlton, C.M., Thompson, B.T., and Mizaikoff, B. (2005) Hollow waveguide infrared spectroscopy and sensing. In: Wolfbeis, O.S. (Ed.) *Frontiers in Chemical Sensors*. Springer, p. 133.

Davies, J.T. and Rideal, E.K. (1963) *Interfacial Phenomena*. Academic Press, Chapter 2.

Demas, J.N., DeGraaf, B.A., and Coleman, P.B. (1999) *Anal. Chem*. 793.

Faist, J., Capasso, F., Sirtori, C., Sivco, D.L., Hutchinson, A.L., and Cho, A.Y. (1994) *Science* 264, 553–555.

Giuliani, J.F., Wohltjen, H., and Jarvis, N.L. (1983) *Opt. Lett*. 8, 54.

Hecht, H.G. (1983) *Appl. Spectroscopy* 37, 348.

Hlady, V., Wagenen, R.A., Andrade, J.D. (1985). In: J.D. Andrade (Ed.) *Surface and Interfacial Aspects of Biomedical Polymers Vol. 2*. Plenum Press, New York.

Hollas, J.M. (2004) *Modern Spectroscopy*. Wiley.

Homola, J. (2004) In: Wolfbeis, O.S. (Ed.) *Optical Sensors*. Springer.

Homola, J. (2008) *Chem. Rev*. 108, 462–493

Janata, J. (1987) *Anal. Chem*. 59, 1351.

Janata, J. (1992) *Anal. Chem*. 64, 921A.

Kraft, M., Jakusch, M., Karlowatz, M., Katzir, A., and Mizaikoff, B. (2003) *Appl. Spectr*. 57, 591.

Kroneis, H.W. and Marsoner, H.J. (1983) *Sens. Act*. 4, 587.

Kubelka, P. and Munk, F. (1931) *Z. Tech. Phys*. 12, 593.

Kuswandi, B., Andres, R., and Narayanaswamy, R. (2001) *Analyst* 126, 1469.

Lechuga, I.M., Prieto, F., and Sepulveda, B. (2004) In: O.S. Wolbeis (Ed.) *Optical Sensors*. Springer.

Lee, M. and Walt, D.R. (2000) *Anal. Biochem*. 282, 142.

Manley, G.A., Klump, G.M., Köppl, C., Fastl, H., and Oeckinghaus, H. (1995) *Advances in Hearing Research*. World Scientific.

McDonagh, C., Burke, C.S., and MacCraith, B.D. (2008) *Chem. Rev*. 108, 400–422

Melling, P.J., and Thomson, M. (2002) *Handbook of Vibrational Spetcroscopy*, Chalmers, J.M. and Griffith, P.R. Eds., Wiley.

Michael, K.L., Taylor, L.C., Schultz, S.L., and Walt, D.R. (1998) *Anal. Chem*. 70, 1242.

Mizaikoff, B. (1995) *Sens. Act. B* 29, 58.

Mizaikoff, B. (2003) *Anal. Chem*. 258A.

Narayanaswamy, R. and Wolfbeis, O.S. (2004) *Optical Sensors*. Springer.

Newby, K., Reichert, W.M., Andrade, J.D., and Benner, R.E. (1984) *Appl. Opt*. 23, 1812.

Okamoto, K. (2000) *Fundamentals of Optical Waveguides (Optics and Photonics)*. Academic Press.

Opitz, N. and Lubbers, D.W. (1983) *Sens. Act*. 4, 473.

Potje-Kamloth, K. (2008) *Chem. Rev*. 108, 367–399.

Rakow, N.A. and Suslick, K.S. (2000) *Nature* 406, 710.

Rosenzweig, Z. and Kopelman, R. (1996) *Sens. Act. B* 36, 475.

Sapsford, K.E. and Ligler, F.S. (2004) In O.S. Wolfbeis (Ed.) *Optical Sensors*. Springer.

Schaffar, B.P.H. and Wolfbeis, O.S. (1988) *Analyst* 113, 217.

Seker, F., Meeker, K., Kuech, T.F., and Ellis, A.B. (2000) *Chem. Rev*. 100, 2505.

Skoog, D.A., Holler, F., and Nieman, T.A. (1998) *Principles of Instrumental Analysis*, 5th edn. Harcourt Brace College Publishers.

Suslick, K.S. (2004) MRS Bulletin No. 10, 720.

Wang, K., Seiler, K., Morf, W.E., Spichiger, U., Simon, W., Lindner, E., and Pungor, E. (1990) *Anal. Sci*. 6, 715.

Wolfbeis, O.S. (2004) Optical technology until the year 2000: an historical overview. In: O.S. Wolfbeis (Ed.), *Optical Sensors*, Springer.

Wolfbeis, O.S. and Offenbacher, H. (1986) *Sens. Act*. 9, 85.

Wolfbeis, O.S. and Schaffar, B.P.H. (1987) *Anal. Chim. Acta*. 198, 1.

Wolfbeis, O.S., Posch, H.E., and Kroneis, H.W. (1985) *Anal. Chem*. 57, 2556.

Yin, T.C.T. and Chan, J.C.K. (1988) Neural mechanisms underlie interaural time sensitivity to tones and noise. In: W.E. Gall, G.M. Edelman, and W.M. Cowans (Eds.), *Auditory Function: Neurobiological Bases of Hearing*. Wiley, pp. 385–430.

Zhang, C. and Suslick, K.S. (2005), *JACS* 127, 11548.

Chapter 10
Multivariate Sensing

10.1 Introduction

The number and types of sensors that are available for most sensing situations are very large; the computational capacity is abundant and cheap. Miniaturization is a strong trend in the chemical sensor field. The stage is then set for extraction of information from data by means of computational multivariate analysis.

The two terms have different meaning. By *data* we mean raw output from the sensor, usually in the form of an electrical signal. In a well-behaved individual sensor, the relationship between the output signal and the concentration of a specific analyte is defined and reproducible. This is what we have learned so far from the discussion of the principles of the individual sensors in the preceding chapters.

But is this what we always want? Very often, we may be interested in information that lies beyond the bare knowledge of concentrations of individual components that sensors can give us, but involves experience, visual information, and other cognitive inputs. In processing information from multiple inputs, we must construct a model in which the concentrations of the components, their ratios, and even their kinetics play a role. Thus, an experienced wine taster can recognize not only the type of wine and the region of its origin, but also the vintage. A trained dog can track a specific smell in the woods. The male gypsy moth can find the female moth over a large distance. Undoubtedly, the information that these examples illustrate is contained in the concentrations, often extremely low concentrations of specific chemicals that are detected by specific olfactory sensors. The signals ("data") from these sensors are then converted by the brain to "information." The result of such an exercise is then: for example, Bordeaux, "Chateau Sissan 2001," in the case of the wine-taster, the dog finding the criminal hiding in the bush, and reproduction or death, as the case may be, for the moth. In each case, an elementary intricate act of sensing has been performed, yet the explicit knowledge of the concentrations of the analytes has not been consciously recognized. The data reduction occurred through some hidden model embedded in the biological object, by training (wine-taster or dog) or evolution (in the case of the moth).

No wonder that these feats of sensing have captured the imagination of sensor scientists and engineers and various engineered "electronic noses" and "electronic tongues" have started to appear since the early 1980s. This is now a very active, often oversold, but rapidly growing area of chemical sensing. In this process, the raw data acquisition is just as important as the data reduction that approximates the desired information, by comparing the defined model with the reality. The data processing is the subject of statistical data analysis methods commonly known as chemometrics.

A word of caution is necessary at the onset. The correct information cannot be obtained if the raw data from the sensors are false. Likewise, an incorrect application of a chemometric technique can lead to misinformation. Thus, the two parts of the information acquisition process are complementary, but can never substitute for each other.

10.2 Higher-Order Sensors

The concept of "order" applies across the analytical field (recall the discussion of kinetics in Chapter 2). Order is also applied in classifying chemical sensors. When only one physical parameter constitutes the output of the sensor and is correlated with concentration, we call it a first-order sensor. An example is optical sensing of a component at one fixed wavelength. The concentration of the unknown sample is then obtained from the calibration curve (Fig. 10.1a) against absorbance, or by a standard addition method. For nonlinear sensors it is possible to use a "linearization" function f.

$$\Re = f(C) \tag{10.1a}$$

For example, the output of a glass electrode (in mV) plotted against the antilog of activity of hydrogen ion yields a linear pH scale. It is the simplest form of performing analysis. This simplicity comes with a price, however. If the sample is contaminated by an unknown impurity, or if the response function $\Re$ changes for whatever reason, an undetectable error accrues. Therefore, the first-order analysis relies on the invariability of the experimental conditions.

$$\Re = f(C, \lambda) \tag{10.1b}$$

The analogy between spectrophotometric analysis and chemical sensing at multiple wavelengths leads to an even stronger case of third-order analysis. When the spectrum is recorded at the output of a chromatographic column, the components are resolved according to their chromatographic retention times, t_R. The response (Fig. 10.1c) can be shown as

$$\Re = f(C, \lambda, t_R) \tag{10.1c}$$

In this case, multiple species can be quantified and the baseline drift can be mathematically corrected. The third-order sensors thus belong to the category
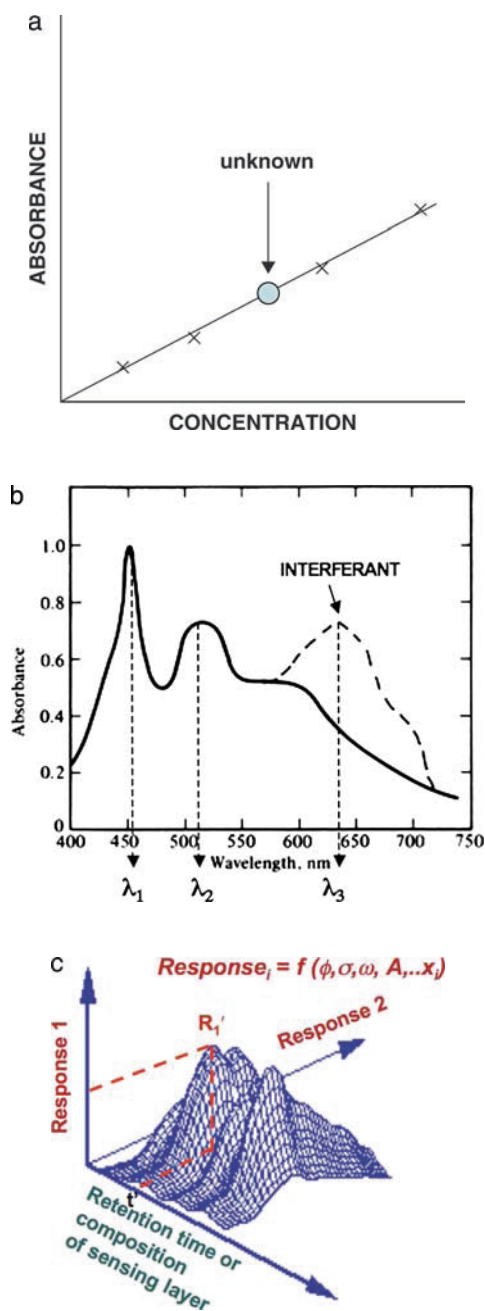
Fig. 10.1 (**a**) First-order chemical sensor in which absorbance is uniquely related to concentration by calibration curve. (**b**) Second-order sensor in which absorbance is shown as a function of wavelength λ. Interferant is easily identified in the spectrum. (**c**) Third-order sensor yielding information in 3-D space. The red dashed line shows conversion of third-order sensor to second-order sensor when the value of response R_1 is obtained at a fixed retention time t'

of "hyphenated analytical techniques," such as GC-MS, HPLC-FTIR, MS-MS, multidimensional chromatographies, and so on.

It is always possible to reduce the order of an analytical technique. This is indicated in Fig. 10.1c by two primed, dashed lines. A good example is GC-MS, where the MS is operated in total ion count mode, only as a chromatographic detector.

Although less common, some third-order chemical sensors have found significant applications not only in sensing but also in research. One such example is Electrochemical Quartz Crystal Microbalance (EQCM). With EQCM, an electrochemical experiment can be performed in its inherently large experimental space, that is, various electrochemical waveforms, impedance analysis, gating, and different mass loading. As the dimensionality of the experiment is increased, so is its information content.

Let us illustrate the benefits of higher order on a concrete analytical example: measurements of concentration of Mg^{2+} with an ISE and with an optical sensor. After linearization of the potentiometric signal, the two experiments can be displayed as a bilinear plot (Fig. 10.2). Contained in this plot is an "unusual" sample point **S**, which clearly falls out of the linear correlation because it lies outside the statistically acceptable 3σ noise level. This outlier is an indication of the presence of an interferant. Its presence is clearly identified in this bilinear plot from combined ISE and optical measurement, although it would be undetected in a first-order sensor alone.

Nevertheless, we still cannot identify which sensor is correct. For that, we need to go to the third-order level by performing, for example, a preseparation. If the two sensors are placed at the outlet of a chromatographic column, the signal for the "pure" sample is shown in Fig. 10.3a and the "contaminated" sample is shown in Fig. 10.3b. There, the retention time for pure standards t_{RS} is different from the retention time for the interferant t_{RI}, which strongly affects the response of the ISE but does not affect the optical sensor output to a significant degree. One such interference could be, for example, a different ionic strength of the "unusual"
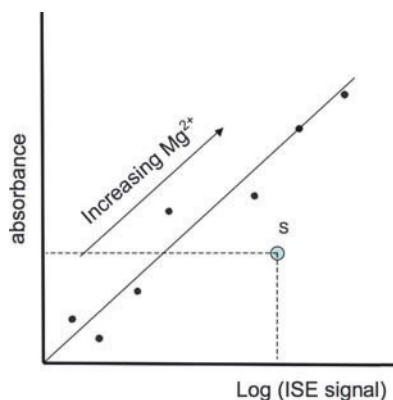
Fig. 10.2 Bilinear plot obtained from second-order sensor. In this example, linearized ISE response is plotted against absorbance from fiberoptic sensor. Point **S** represents "an outlier"
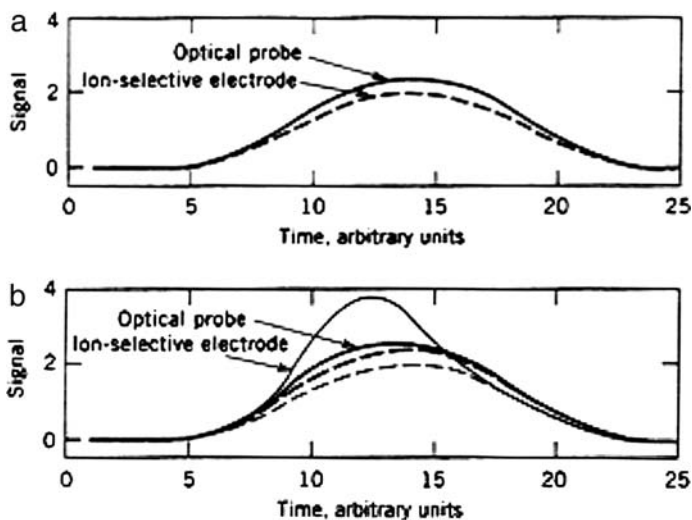
Fig. 10.3 Identification and quantification of the "interferant" by a third-order sensor in which optical, ion-selective electrode, and chromatographic data are combined: (**a**) pure standards; (**b**) "contaminated" sample

sample. Therefore, third-order sensors can identify interferants and quantify the sample.

10.3 Sensing Arrays

Another way to expand the information acquisition space is through the use of chemical sensing arrays. In this case, the same transduction modes apply but the number of selective layers is increased. In certain ways this approach is similar to the experiment in which the output of the chromatographic column is monitored by a spectrophotometer (Fig. 10.1c). Each selective layer represents one channel of sensing information, just as specific retention times identify specific components in a given mixture. Ideally, the layers in the array are designed in such a way that they respond uniquely to only one species of interest; that is, they are orthogonal. This requirement is rarely satisfied and some degree of cross-selectivity always exists between individual sensors. It would correspond to the overlap of the peaks in the chromatographic analogy.

Here also lies one major difference between a sensor array and a separation experiment. In the latter, the advancing front of the already separated sample always encounters the pristine stationary column phase and interacts accordingly, whereas in a chemical sensor array all components of the sample interact with the sensing layers at once. The direct sensing and the separation experiments are not directly comparable but are complementary. In summary, increasing the order of the sensor

or employing sensing arrays, or combining both strategies, makes available a very large amount of data. The conversion of raw data to information is the subject of statistical methods collectively known as chemometrics.

Information obtained from the array is much more than the sum of information obtained from the individual sensors. This additional information accrues from analyzing and understanding the group behavior of chemical sensors. There are many excellent texts covering this topic in detail. Reviews by Hierlemann et al. (1996) and Jurs et al. (2000) are particularly recommended. A large section of this chapter follows, and in some spots paraphrases (in abbreviated form), those two reviews.

10.4 Multivariate Analysis

In this section we examine a few representative samples of multilinear techniques.

10.4.1 Preprocessing of Input Data

Raw signals from chemical sensors are rarely suitable for direct multivariate analysis. Some form of signal conditioning is always necessary before the input matrix is composed. Examples of preprocessing techniques used in the static and in the dynamic mode of multicomponent analysis are summarized in Table 10.1. They can be used as such or in combination. In higher-order sensors, where different transduction modes are used, the homogeneity of the input matrix is important. Thus, the matrix must contain data that are comparable in dimensions and that are commensurate.

Table 10.1 Examples of static and dynamic modes of sensing arrays and their uses

Static mode	Use
Subtraction of reference	Differential measurements
Linear drift subtraction	Baseline correction
Normalization	Relative signals
Averaging	Signal quality; redundancy
Linearization	Transform functions, e.g., antilog

Dynamic mode	Use
Steady-state extrapolation	Speeding up of response
Fourier analysis	Spatial and temporal information
Gating and signal filtering	Enhancement of selectivity
Transient signals	Increasing order of measurement
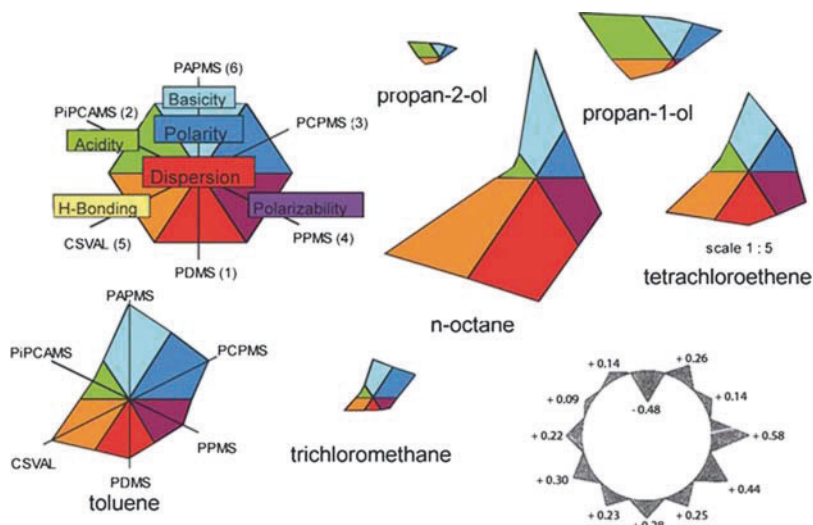
Fig. 10.4 Examples for polar plots for various vapors according to the LSER interaction model (*top left*). A quantitative value is plotted on each "polar" coordinate. Change of "shape" of the polar plot is shown for different analytes. At the bottom right is the "offset" polar plot that accommodates bipolar, that is, $+/-$, responses (adapted from Hierlemann et al., 1996)

10.4.2 Data Display

Graphic display of raw data can help orient the analyst and can even facilitate the selection of the most appropriate data processing technique. Various graphing software can be used for this purpose. Radial polar plots have been particularly popular because they allow classification according to the physically significant coordinates. For example, in the interaction of analytes with the selective materials classified by the LSER relationship (Chapter 2), the axes can be labeled according to the principal interaction components, which are then evaluated from the experimental data fitting and plotted in form of the polar plot. In that example, the data in the plot are monopolar. If the data are bipolar, an offset polar plot with an arbitrarily selected offset value can be used (Fig. 10.4). The value of this simple data processing is based on the ability of the human eye to recognize and classify complex geometrical shapes.

10.4.3 Data Processing

The number of methods of data analysis is so large and the choices are sometimes so confusing, that a roadmap is needed for orientation (Fig. 10.5). At the top of this map are the model-based techniques that require specific and exact correlation between physical parameters and concentration, therefore calibration. At the bottom are the techniques that are model-free. They search for similarities in the response
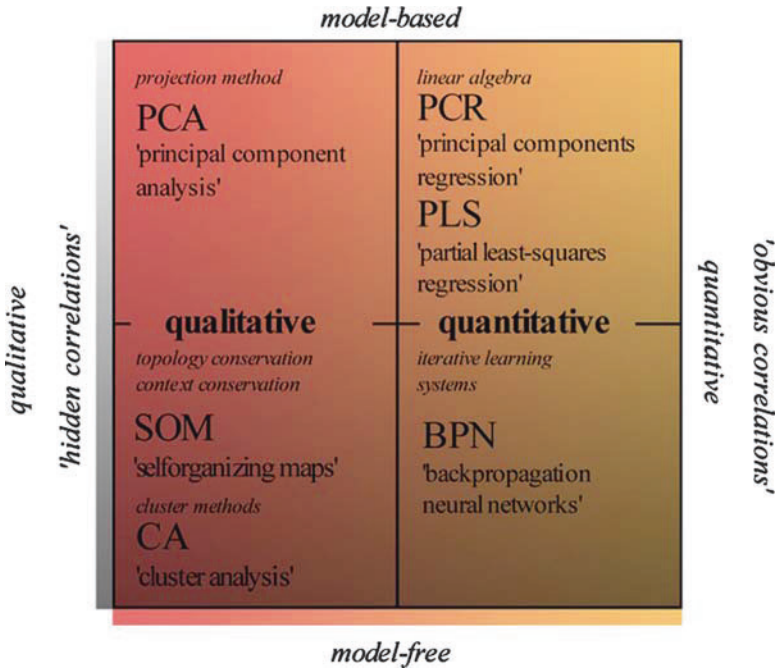
Fig. 10.5 Roadmap of various common techniques used in multivariate analysis as discussed in this chapter (adapted from Hierlemann et al., 1996)

without attribution of such response to any specific physical parameter. The left-to-right direction represents the transition from unsupervised to supervised algorithms. Unsupervised algorithms require only the values of the signals, whereas supervised algorithms rely on the explicit relationship between the signal and the concentration. The latter can be used for either qualitative or for quantitative evaluation.

There are some conventions and terminology that are common to all statistical data evaluation approaches. As the first step the input matrix X is organized such that the independent variables (compositions) are arranged in rows from 1 to t and the preprocessed signals from sensors 1 to m are entered in the columns.

$$\underline{\underline{X}}_{t \times m} = \overbrace{\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{t1} & x_{t2} & \cdots & x_{tm} \end{pmatrix}}^{Variable\,1\ \ Variable\,2\ \ ...\ Variable\,m} \underbrace{\left.\begin{matrix} Mixture\ 1 \\ Mixture\ 2 \\ \vdots \\ Mixture\ t \end{matrix}\right\} Objects}_{Variables\ (Sensor\ Signals)} \qquad (10.2)$$

The input matrix can be transformed in such a way that unique vectors can be defined independently of each other. Such vectors then describe the feature space.

This process is called orthogonalization. From the sensing array point of view, it is desirable to have sensors and responses that have features that are maximally independent of each other. Within the matrix itself can be found nonzero vectors whose direction is not changed by such a transformation. These are so-called eigenvectors (meaning approximately "characteristic vectors"). In the transformation, the value of the eigenvector vector can, however, change by a factor, which is called the eigenvalue. The set of eigenvectors of the same eigenvalue defines the eigenspace.

The input matrix from (10.2) can be transformed to a new eigenspace.

$$\underline{\underline{X}}_{t \times m} = \underline{\underline{U}}_{t \times m} \underline{\underline{S}}_{m \times m} \underline{\underline{V}}^T_{m \times m}$$

$$= (\underline{u}_1, \quad \underline{u}_2, \quad \cdots \quad \underline{u}_m) \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_m \end{pmatrix} (\underline{v}_1, \quad \underline{v}_2, \quad \cdots \quad \underline{v}_m)^T \qquad (10.3)$$

Here the matrix S_{mxn} denotes the diagonal matrix of singular values and the matrix U_{rxm} is the matrix of the "left eigenvectors." The matrix V_{mxm} contains the "right eigenvectors," which are uncorrelated linear combinations of the original input variables. The singular values are positive square roots of the eigenvalues of the covariant matrix, arranged in decreasing order (e.g., $s_1 > s_2 > \dots > s_m$). The mathematical rank of matrix X then equals the number of these nonzero s-values. The simplest and most popular model-based techniques assume a linear relationship between the sensor signal and concentration of the analyte. This greatly simplifies the calibration by building concentration-signal pairs from known regression parameters.

10.4.3.1 Principal Component Analysis

The simplest and most widely used chemometric technique is Principal Component Analysis (PCA). Its objective is to accomplish orthogonal projection and in that process identify the minimum number of sensors yielding the maximum amount of information. It removes redundancies from the data and therefore can be called a true data reduction tool. In the PCA terminology, the eigenvectors have the meaning of Principal Components (PC) and the most influential values of the principal component are called primary components. Another term is the loading of a variable i with respect to a $\mathbf{PC_f}$.

$$\text{Loading}\ (i, \mathbf{PC_f}) = s_f \cdot v_{if} \qquad (10.4)$$

High loading of $\mathbf{PC_f}$ indicates that this vector is aligned close to the original data values; that is, the transformation to the feature space defined by the new principal components matches the important (perhaps the most important) trend in the raw data. Conversely, low loading means that the PC does not match a significant trend in the data. Typically, 2–3 PCs can characterize most experimental datasets. This then allows a 2-D or a 3-D graphical representation of the results, as shown in Fig. 10.6. As powerful as it is, the PCA fails in cases where the individual sensors in the
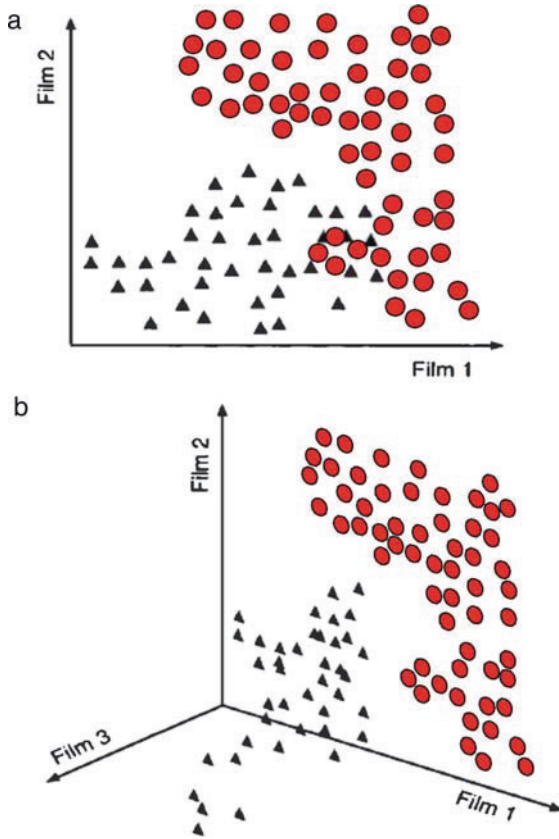
Fig. 10.6 Principal component analysis (PCA): (**a**) 2-D PCA in which two datasets are partially overlapping; (**b**) 3-D PCA in which addition of the orthogonal third component resolves the overlap (Ricco et al., 1998)

array behave nonideally. This could mean hysteresis, sample-dependent or time-dependent deviation from linearity, or nonadditivity. In short, the safe application of PCA requires well-behaved sensors. Another problem lies in the visual evaluation of the PC plots that may be affected by subjective error. An example of it can be seen in Fig. 10.6a where overlap of two data clusters could present a problem. It can be eliminated by adding the third PC (Ricco et al., 1998).

10.4.3.2 Principal Component Regression

This is a calibration technique (i.e., supervised) that yields quantitative information about the components in the sample according to the relationship

$$C = XP \tag{10.5}$$

Note that this equation describes the relationship between concentration C of the component and the sensor response X. It is purposely written "backwards" by comparison with the usual notation used with linear sensors (e.g., optical, amperometric, etc.) discussed earlier. This convention helps to define P as the matrix of regression coefficients.

For a well-behaved sensor array, only a small subset k of n available PCs is sufficient to characterize the matrix. Once again, Principal Component Regression (PCR) is a data reduction tool. The robustness of the selection of k can be tested by cross-validation in which case data subsets are randomly selected and the error matrix $\boldsymbol{H_{nxn}}$ is calculated.

10.4.3.3 Partial Least Squares

Another solution of (10.5) called bidiagonalization is used in Partial Least Squares (PLS). Here the original input matrix X is linked to two orthogonal matrices O and W by a bidiagonal matrix L.

$$\underline{\underline{X}}_{t \times m} = \underline{\underline{O}}_{t \times m} \underline{\underline{L}}_{m \times m} \underline{\underline{W}}^{T}_{m \times m}$$

$$= (\underline{o}_1, \quad \underline{o}_2, \quad \cdots \quad \underline{o}_m) \begin{pmatrix} l_{11} & l_{12} & 0 & \dots & 0 \\ 0 & l_{22} & l_{23} & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & l_{m-1\,m-1} & l_{m-1\,m} \\ 0 & 0 & \dots & \dots & l_{mm} \end{pmatrix} (\underline{w}_1, \quad \underline{w}_2, \quad \cdots \quad \underline{w}_m)^{T}$$

$$(10.6)$$

In contrast with PCR, the PLS factors (vectors w) are not independent of each other.

The selection of PLS factors is therefore simplified because they are a priori ranked according to their value.

The validity of the parameterized techniques discussed thus far relies on the linear and stable relationship between concentration and response of individual sensors. With this caveat in mind, these data reduction and evaluation tools can also be used for optimization of arrays in terms of numbers and orthogonality.

Moving down on the multivariate data analysis roadmap (Fig. 10.5) we come to the nonparameterizing approaches. They do not require any model and do not depend on the linearity or statistics of the concentration/response relationship. In this model-free domain are two architectures that are discussed as examples: *unsupervised*, Self-organizing Maps (SOMs, also known as Kohonen maps) and *supervised* Artificial Neural Networks (ANNs, sometimes called "perceptrons"). They both utilize iterative adjustment of gains on the individual nodes in the learning process. Because of their semantic similarity to biological systems, terms such as *neurons* and *synapses* are often used. Such an analogy can be both inspiring and misleading, but it is already an entrenched part of the jargon. This form of data

processing is extremely active not only in chemical sensing, but generally, in other applications of artificial intelligence.

10.4.3.4 Self-organizing Maps

The principle of Self-organizing Maps (SOMs) is outlined in Fig. (10.7). In the first iteration, the training output from the m sensor array is entered into the input layers X_1 and X_2. From there, it passes through the weighing nodes W, which have weighing factors randomly selected between 1 and $+1$. They are connected to each element of the output layer, creating a 2-D output space. In the next iteration, the "neighborhood relations" are prescribed to the weighing elements by a Euclidian algorithm that quantitatively characterizes the similarity between the input and the output points. The objective of the iteration is to minimize the distance between input and output elements, in other words to promote the clustering of the elements to match the values defined by the training set. The commonly used Kohonen algorithm, resulting in the "neighboring" effect, can use either a "bell-shaped" Gaussian or a "Mexican hat" function. The equation for the Gaussian filter is

$$\Lambda(\delta\varepsilon, \beta\gamma) = \exp\left[\frac{-\left|\underline{r}_{\delta\varepsilon} - \underline{r}_{\beta\gamma}\right|^2}{2\sigma^2}\right] \qquad (10.7)$$

where $\left|\underline{r}_{\delta\varepsilon} - \underline{r}_{\beta\gamma l}\right|$ is the distance of the output element $\delta\varepsilon$ from the "winner" (i.e., true value) $\beta\gamma$. A major advantage of this approach is that the training is done efficiently and the set often converges rapidly. Similar to SOM is another clustering technique, Adaptive Resonance Theory (ART). The significant difference between SOM and ART is that the latter uses nonnormalized inputs, resulting in a 3-D map.
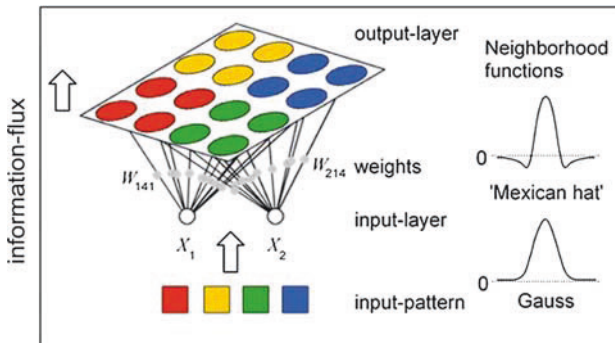
Fig. 10.7 Principle of self-organizing map (Kohonen map) with two typical "neighborhood functions" (Hierlemann et al., 1996). The "information content" increases from the "bottom up" during the self-assembly

10.4.3.5 Neural Networks

Known also as artificial neural nets (ANNs, or perceptrons), they represent a very large group of complex data processing techniques that have found applications in situations where a very large number of complex inputs need to be processed rapidly, such as economic forecasting, psychology, navigation systems, and security, among others. They employ a fundamentally different computational approach from ordinary digital computers, based on the serial, von Neumann processing paradigm. Neural networks are more akin to brain functions where data processing takes place in a massively parallel architecture. In principle, Neural Networks (NNs) cannot do anything that could not be done by serial digital computing, but they can do computations that would otherwise be difficult to do. There is one feature of NNs that is particularly attractive, albeit somewhat surprising. They can perform calculations without prior knowledge of the model (i.e., algorithm). In fact, NNs can be used to create models. In the world of manmade machines, because of their parallel mode of operation they are closer to analog than to digital computers. Not surprisingly, NN chips dedicated to specific computational tasks have been fabricated and are commercially available.

There have been many books and reviews written on the subject of NN and parallel computing. Only a token one is listed here, for those who need a traditional book reference (Haykin, 1999). It will probably be obsolete before this book is published. Otherwise, a wealth of up-to-date information is always available on the Internet where a "neural networks" entry produces an avalanche of information. Both lead articles cited for Chapter 10 (Hierlemann et al., 1996) and (Jurs et al., 2000) discuss their applications in the context of chemical and biological sensing.

The architecture of a common NN is shown in Fig. 10.8. The design depends on the types of sensor responses, on their dynamic range, drift, and so on. In short, it depends on all the complexities of the transfer functions of different types of sensors. Once again there is an input layer containing m input elements. It is massively interconnected to the n nodes of the next hidden layer χ_n at which the weighing factors W_n operate on the signal. There can be more than one hidden layer, if necessary. The connection to the output layer Θ_o has the form of nonlinear transfer function f_{hid}; for example,

$$f_{hid}(x) = f_{out}(x) = \tanh(\beta x) \qquad (10.8)$$

This is useful for pattern recognition. The parameter β influences the nonlinearity of the hidden layer transfer function. The training of the NN is then based on the feedforward and backpropagation algorithm, in which the weighing factors W_n are adjusted during the NN learning in order to minimize the difference between the desired output D and the actual output Y.

$$\Delta W_i = \eta(D - Y)x_i \qquad (10.9)$$

The symbol η stands for the adjustable learning rate. It determines how much time the NN requires to converge on the correct result (if ever!). The fact that there are a great many adjustable parameters, functions, and architecture modifications makes
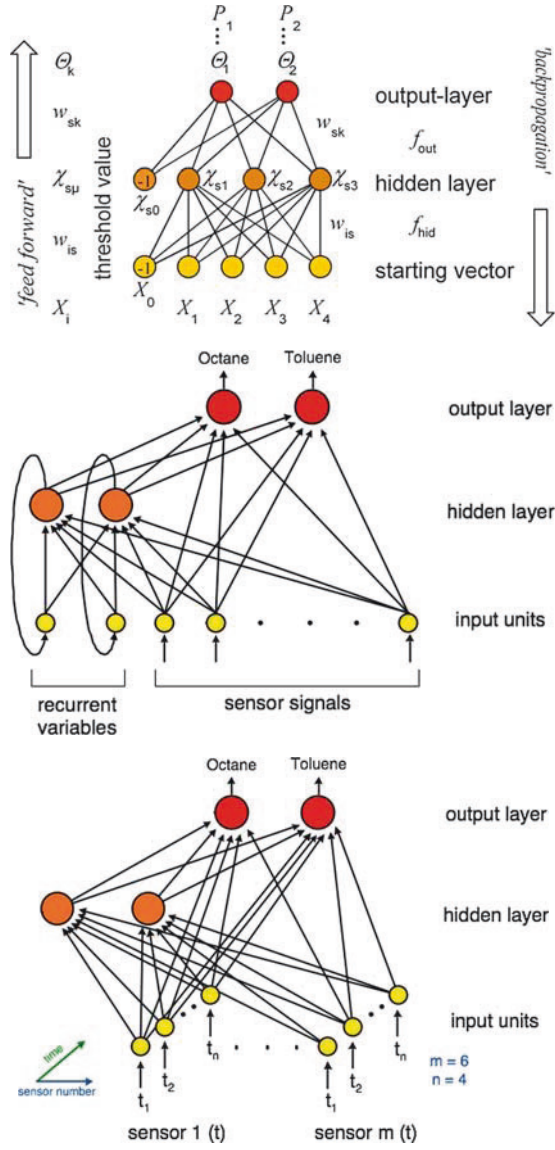
Fig. 10.8 (**a**) Example of common neural net (perceptron) architecture. Here "one hidden layer" Neural Networks (NNs) is shown (Hierlemann et al., 1996). (**b**) A more sophisticated "recurrent neural network" utilizing adjustable feedback through recurrent variables. (**c**) Time-delayed neural network in which "time" has been utilized as an experimental variable

the NN adaptable to essentially any application. However, there is a price associated with this flexibility, such as long learning of the NN, inability to converge on a correct solution, and possibility of artifacts due to "overlearning." This becomes a problem particularly if an unknown component, which has not been present in the training set, unexpectedly appears in the real sample.

10.4.3.6 Cluster Analysis

Grouping of samples by their similarity is another important sensing task. Most techniques discussed above will yield clusters of similar responses. The similarity is computed as distance d between two samples x_{ik} and x_{jj}, which in p-dimensional space is given by the formula

$$d = \left(\sum_{k=1}^{N} (x_{ik} - x_{jk})^p \right)^{1/p} \tag{10.10}$$

N represents the number of sensors in the array. For $p = 2$, the distance in (10.10) is Euclidian. The protocol is relatively simple. The distance matrix is created from the datapoints and scanned for the smallest values that are then arranged and displayed in the form of a dendrogram (Fig. 10.9; Suslick, 2004) in which the dissimilarity is plotted on the horizontal axis. In a dendrogram, each horizontal line segment represents the distance—that is, the similarity—between samples. Thus, if we want
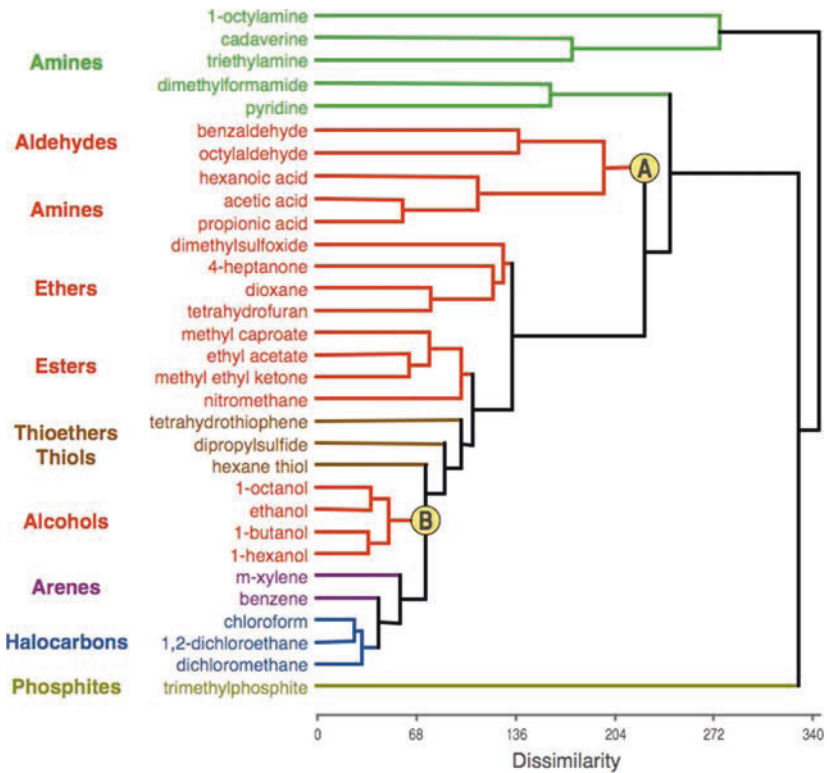
Fig. 10.9 Example of dendrogram. The "dissimilarity" between compounds is shown as Euclidian distance on the horizontal axis (Suslick, 2004)

to assess the similarity between, for example, hexanoic acid and octanol (Fig. 10.9), we have to travel the path shown in red, to point A (dissimilarity score 230). In this case, the dissimilarity is represented by the point A which is farthest to the right on the *x*-axis. On the other hand, similarity between octanol and chloroform (green path) would bring us to point B (dissimilarity score 85). Therefore, octanol and chloroform behave more similarly on this array. The dendrogram thus becomes a simple and powerful tool for assessing the similarity in p-dimensional feature space. However, it is important to realize that this is similarity between the response of the sensor to the set of chemicals and not necessarily the similarity between their chemical nature.

10.4.3.7 Visual–Empirical Region-of-Influence

This technique makes use of the innate human cognition to perform clustering in multidimensional space (Osbourn and Martinez, 1995). It is unsupervised and model-free, therefore requires from the user only the data input. A special "mask" mathematically defines the "visual region of influence." Its shape is based on human visual perception, taking advantage of the human brain to recognize and cluster objects (Fig. 10.10a). Its properties are as follows. Two points in space are clustered only if no other point lies within area of the "mask" which thus defines the exclusion region. In this way, an n-dimensional problem is reduced to set of *n* two-dimensional problems.

Although the shape of the Visual–Empirical Region-of-Influence (VERI) mask is invariable, it size scales automatically according to the properties of the cluster (Fig. 10.10b). The VERI algorithm requires preprocessing of the data and for that purpose PCA or PCR preprocessing is routinely used. It is relatively immune to the presence of unknowns, and nonlinearity and nonadditivity of sensor responses (Osbourn et al., 1998). It has been used successfully to determine the optimum
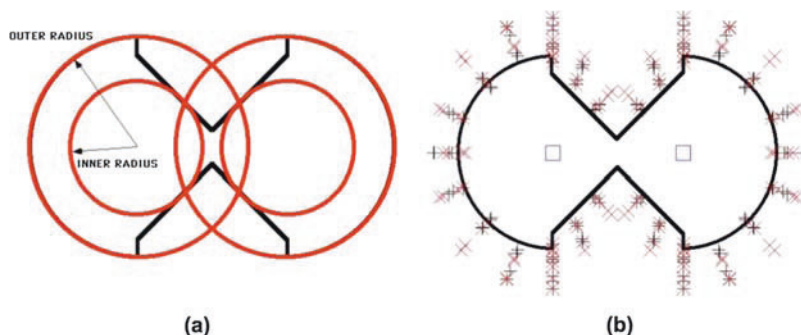
Fig. 10.10 "Rejection mask" used in Visual Empirical Region of Influence (VERI). The shape of the mask is mathematically derived from the visual cognitive factors: (**a**) geometrical scaling relationships used in the derivation of the "shape"; (**b**) final mask which defines the "exclusion" region, that is, objects that "do not belong to the group" (Osbourn and Martinez, 1995)

number of sensors in the array. The exact answer to that question depends on the degree of orthogonality of the sensors in the array, but it has been shown that for polymer-coated SAW sensors the optimum number is between 8 and 12. Increasing the size of the array is counterproductive because it increases the information noise without increasing the information content. It actually degrades the performance of the array.

10.4.3.8 Hybrid Approaches

The preprocessing mentioned above is just one example of the hybridization multivariate technique. In essence, several multivariate analytical techniques are used in various combinations. Kohonen self-organizing maps are frequently used as the first stage in such a processing cascade. It performs the function of "presorting" of the data. The choice of the techniques for the hybrid is dictated by the application and by the behavior of the sensors in the array. An example of the hybrid cascade is shown in Fig. 10.11 where the input stages are unsupervised and perform qualitative sorting whereas the final stages quantify the results.
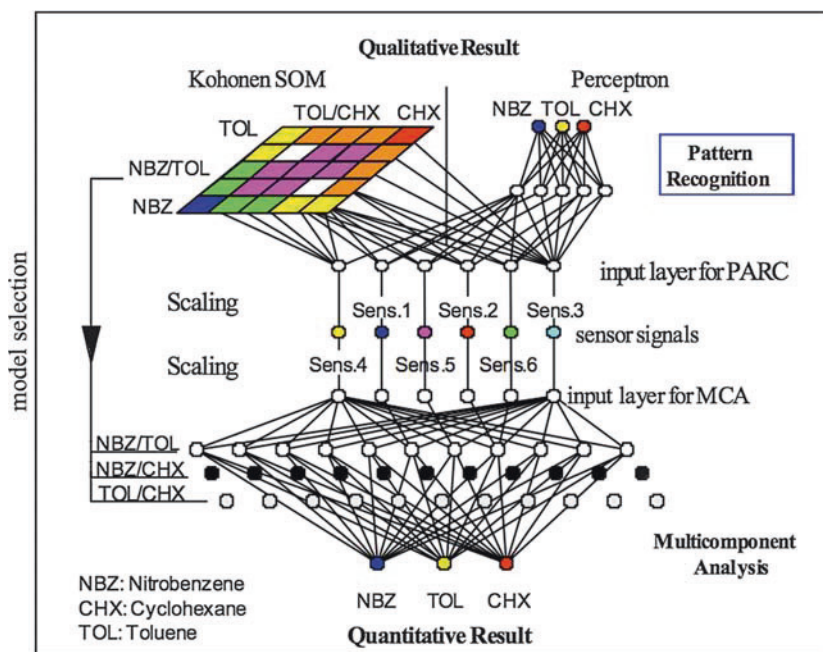
Fig. 10.11 Representation of "hybrid" approach in which Kohonen maps, neural nets, multiple component analysis, and pattern recognition are combined to create a complex data evaluation cascade. Within this cascade supervised (quantitative) and unsupervised (qualitative) techniques are combined (Hierlemann et al., 1996)

10.4.4 Evaluation Dynamic Data

Time is an important parameter in chemical sensing, both where the kinetic selectivity is concerned and also in the most general aspect of response time. Time is a consideration in the operation of individual sensors and in sensing arrays alike. In multivariate analysis, some dynamic characteristic of the sensor such as response time constant or gating (acquisition of data in a given time interval), quenching times, and the like are identified with specific chemicals and then used as one of the parameters in the input data matrix. Most sensor characteristics discussed thus far have been time-invariant, that is, steady-state or equilibrium. Increasing the sensing order by evaluation in the time-variant domain can potentially be a mixed blessing and must be approached with great care.

10.4.4.1 Time as a Parameter

The usual approach in dealing with a monotonic, time-varying signal is to use some form of exponential function with different coefficients for different sensing elements. This is the idea behind Multiexponent Transient Spectroscopy (METS; DiNatale et al., 1992). Although most nonlinear time-dependent responses can be decomposed to n-multiple exponentials or other fitting functions, it is necessary to provide the physical justification for doing so. In other words, the fitting function must bear at least some relationship to the dynamic behavior of the species to which that time parameter is assigned. In chemical sensing, an exponential function usually implies some activated process with an activation energy barrier in the exponent, such as it appears in chemical kinetics. However, when the response is limited, for example, by diffusion or by adsorption kinetics, the dynamic function may have a highly nonexponential time relationship to concentration and simple exponential fitting may result in an information artifact.

The complexity of the dynamic approach is illustrated by transient analysis of an interdigitated CMOS capacitor sensing array consisting of seven sensors, coated with thin polymeric sensing layers and five reference capacitors. The sensing part of the chip was integrated with CMOS circuitry to allow a sampling frequency of 20–40 Hz. The sensing signal was capacitance measured at 1 MHz (Kummer et al., 2006). The analyte vapors were presented to the array as steps in concentration resulting in transient change of the capacitance on a 1–20 s time scale, as the vapors were absorbed by the polymers. The capacitance changes were caused by different physical processes. These were diffusion of the vapor into the polymer interior, swelling of the polymer, and induced changes of the dielectric constants. The first two have their own characteristic time constant. The diffusion coefficient and, to some degree, the polymer swelling depend on the size of the molecule, whereas the dielectric constant is dominated by its dipole moment. Moreover, they are dependent on each other in a rather complex way. Considering these factors, the model of the transfer function has been developed and tested with mixtures of simple alcohols

and water vapor. The normalized array response (Fig. 10.12) was then evaluated by principal least square regression.

When used in the time-invariant mode (i.e., in equilibrium), it is a first-order chemical sensor that can yield qualitative and quantitative information based on the LSER paradigm about composition of the vapor mixtures (Fig. 10.13). By acquiring the data in the transient regime, it becomes a second-order sensor and in addition to the composition, information about diffusion coefficients in different polymers is obtained. This is then the added value. It is possible only because the model describing the capacitance change included diffusion. In spite of the complexity of the response function, a good discrimination and quantification has been obtained.
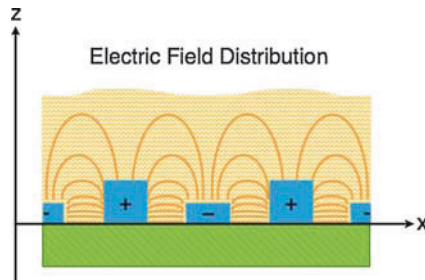
Fig. 10.12 Electric field distribution in the capacitive sensing element utilizing time transients as a parameter (Kummer et al., 2006)
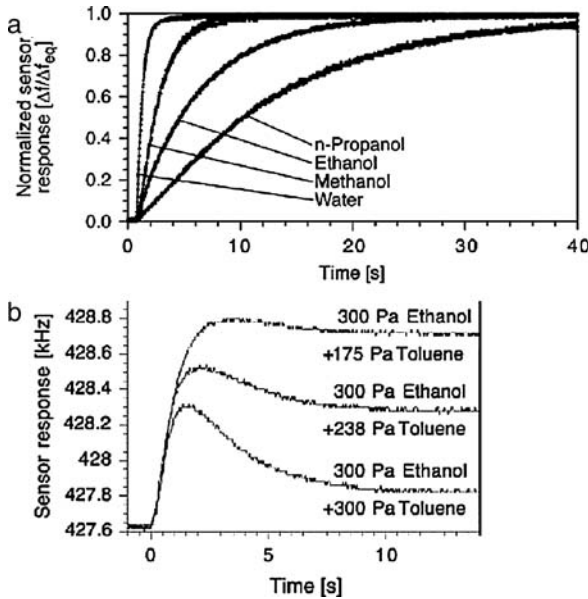
Fig. 10.13 Transient signals obtained from the capacitive sensing element (Fig. 10.12), used for multivariate dynamic analysis (Kummer et al., 2006). Data for (**a**) pure alcohols and (**b**) ethanol/toluene mixtures of different ratios

Another example of a transient array is the set of microelectrodes on which cyclic voltammograms are recorded and a suitable pattern recognition technique is used to analyze it. Clearly, the boundaries of information acquisition can be greatly expanded by the inclusion of time and by careful analysis of the transient signals.

10.4.4.2 Spatial Information Obtained with Sensing Arrays

Thus far, the discussion about sensing arrays revolved around composition of chemical mixtures. In this section, we discuss their use for obtaining information about distribution of chemicals in space. This aspect of chemical sensing is inspired again by biology. Living organisms rely on chemical information for many vital functions, such as search for food, mating, avoiding predators, and so on (Atema 1996). Following traces of chemicals in space is called chemotaxy (Nakamoto et al., 1996). Such a task can be accomplished, in principle, by only one sensor, although arrays of sensors have been used in the design of engineered robots for tracing gases in air. The true benefit of the sensing array becomes apparent when we take advantage of the geometry of the array (Kikas et al., 2000). A spreading chemical in a turbulent stream forms a plume (Fig. 10.14). In this case a red dye has been used and it reveals that such a chemical plume has a certain structure. Thus, the boundaries of the patches of the dye close to the nozzle are sharper than those of the patches that are farther away, downstream. The two "channels" originating from the nozzle
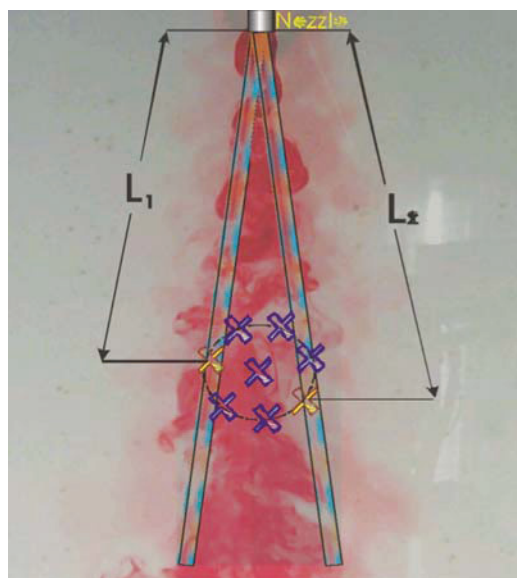
Fig. 10.14 Chemical plume of red dye spreading in turbulent stream. Highlighted sensors from the array are positioned downstream from the nozzle at distances L1 and L2. They are selected for correlation analysis

represent the hypothetical trajectories that the dye has to travel in order to reach the array of nine sensors positioned downstream. The pair of sensors intercepting those channels is highlighted in yellow.

Let us imagine for a moment that an array like this represents an aquatic animal, for example, a crab that has its olfactory organs on its appendages. The spacing between a pair of such sensors is between 10 and 15 cm, depending on the size of the crab. Let us assume that the nozzle represents a source of food and that the chemical information emanating from such a source is common in some way to both channels. Because we are not interested in the chemical analysis, the nature of the sensing mechanism that produces the time-varying signal $S(t)$ is irrelevant. In that case the signal, which is proportional to concentration, can be decomposed to its mean value $<S>$ and its time fluctuating part $s(t)$.

$$S(t) = <S> + s(t) \tag{10.11}$$

The practical relevance of this example becomes apparent when we realize that a physical object, for example, a cylinder with diameter d, is placed in a flowing stream of linear velocity V. It generates a "Kalman vortice street" of eddy currents. The shedding frequency f of the eddies is given by the empirical formula

$$f = \frac{S_t}{Vd} \tag{10.12}$$

The constant S_t has the value of 0.21 for Reynolds numbers in the range 10^2–10^5. Therefore, the nozzle (or the cylinder in the stream) generates a deterministic relationship in the fluctuating signal $S(t)$ in both channels. Let's make another conceptual step and assume that all of the information is contained in the fluctuating part $s(t)$ of (10.11). This may feel counterintuitive because until now we have always utilized the mean value of the signal and regarded the fluctuating part as noise. If we accept this paradigm shift, then we can perform a correlation analysis on $s(t)$, either in the time domain or, equivalently, in the frequency domain. The latter is preferred because it yields spectral information. The first step is to compute the autocorrelation functions R_{XX} and R_{YY} for each channel.

$$R_{XX}(\tau) = \lim_{T \to \infty} \int_0^T x(t)x(t+\tau)\,dt \tag{10.13a}$$

and

$$R_{YY}(\tau) = \lim_{T \to \infty} \int_0^T y(t)y(t+\tau)\,dt \tag{10.13b}$$

This tells us how much energy in the original signal remains after some arbitrary delay time τ. Next we compute the cross-correlation function $R_{XY}(\tau)$, which contains information about energy shared between channel X and channel Y.

$$R_{XY}(\tau) = \lim_{T \to \infty} \int_0^T x(t)y(t+\tau)\,dt \qquad (10.14)$$

Fourier transformation of (10.13) and (10.14) to the frequency domain yields energy spectral densities, specifically

$$S_{XX}(\omega) = F[R_{XX}(\tau)] \qquad (10.15a)$$
$$S_{YY}(\omega) = F[R_{YY}(\tau)] \qquad (10.15b)$$
$$S_{XY}(\omega) = F[R_{XY}(\tau)] \qquad (10.15c)$$

We are interested in the energy that is common to the two signals and define coherence γ^2 as the ratio of cross-energy spectral density $S_{XX}(\omega)$ to the product of autospectral energy densities.

$$\gamma^2 = \frac{|S_{XY}(\omega)|^2}{S_{XX}(\omega) \cdot S_{YY}(\omega)} \qquad (10.16)$$

Coherence is a squared, frequency-dependent correlation coefficient. It has a convenient physical meaning: for signals in channels X and Y that are fully correlated, it is unity. Conversely, for completely uncorrelated signals, it has a value of zero. Thus,

"Uncorrelated".... $\Leftarrow 0 < \gamma^2 < 1 \Rightarrow$ ...."Correlated"

The graphical representation of this protocol is shown schematically in Fig. 10.15. Signals from two amperometric electrodes, representing channel 1 (blue) and channel 2 (red) detect to electroactive species, which is delivered to them with frequency modulation of, for example, 1 Hz. The experiment is performed in the benchtop fluid setup shown in Fig. 10.16. The first interesting observation is the presence of higher harmonics in the coherence spectrum. They arise as the effect of "nonsinusoidal" modulation. A pure sine wave would transform to the frequency domain as a single line. Any other waveform of the same frequency will contain higher harmonics in the spectrum.

A series of experiments has been performed with this "virtual plume." These include the determination of the experimental bandwidth, the effect of dilution on coherence and the dependence on flow rate and distance. What has been found, in summary, was that it is possible to "write-and-read" fundamental frequencies up to 4 Hz and that higher harmonics could be observed up to 10 Hz (Fig. 10.17).

The composite frequencies have been preserved in modulation experiments with two unharmonic frequencies (Fig. 10.18; Kikas et al., 2001). The coherence spectra are relatively immune to the presence of "white noise," as is shown by the red spectrum in Fig. 10.18. No effect on information content was observed with flow rate or dilution up to seven times.

An interesting and perhaps more important result was obtained from the investigation of the distance dependence (Fig. 10.19). The objective of this experiment was to find out how far downstream the modulation was preserved and if some property
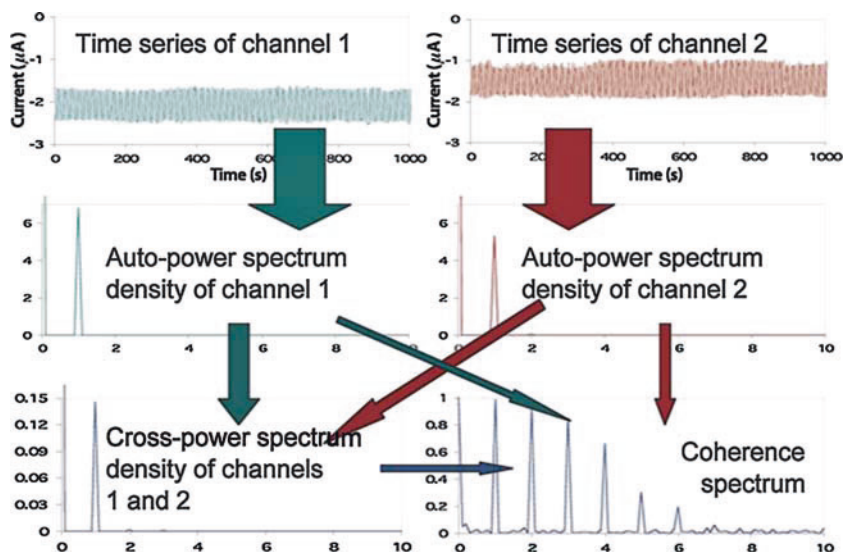
Fig. 10.15 Outline of steps performed in the correlation analysis. In the first step the time-varying signal is Fourier transformed to the frequency domain creating auto- and cross-spectral densities. They are then normalized to form coherence
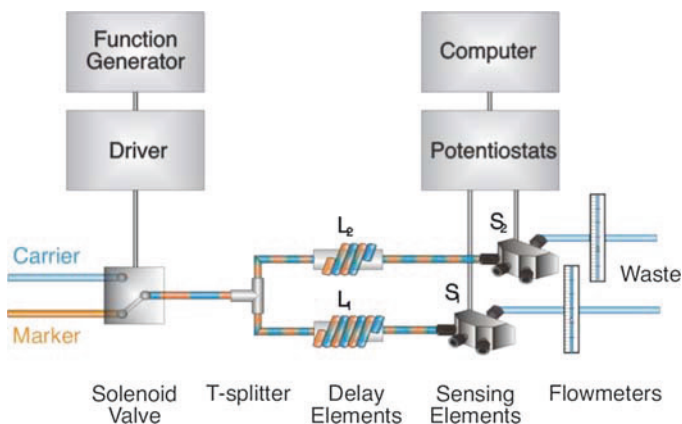
Fig. 10.16 Experimental setup of a "virtual plume" used for verification of the correlation analysis concept. Frequency of the fluctuation of the concentration of the electroactive marker is controlled by the solenoid valve (Kikas et al., 2001a)

of coherence conveyed information about the distance of the array from the source of modulation. The ratios of values of coherence at the fundamental and first harmonic frequency showed "breaks" that correlated well with the distance up to approximately 2 m (Fig. 10.19b). This result confirms the "softening" of the edges of the patches of the chemical rather away from the source as seen in Fig. 10.14. This effect is responsible for the disappearance of the higher harmonics with distance
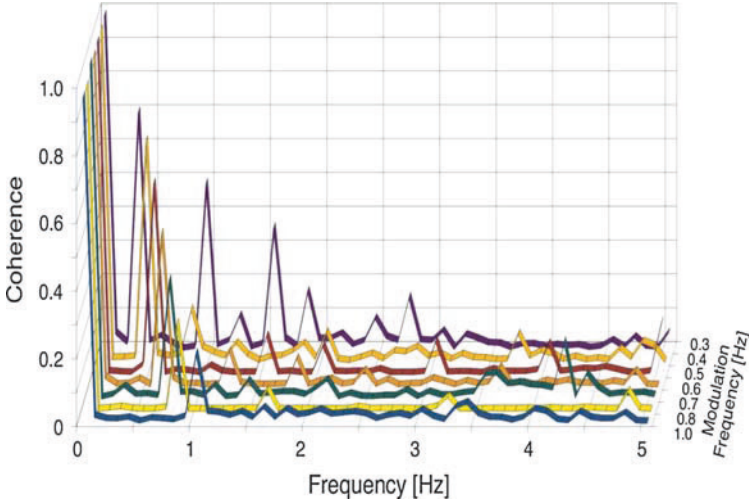
Fig. 10.17 Coherence spectra obtained for modulation frequencies between 03 and 1.0 Hz. Note the faster disappearance of the higher harmonics from the higher-frequency modulation spectrum
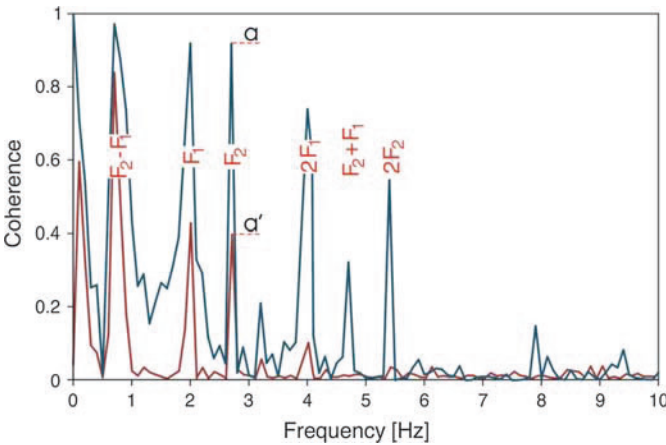
Fig. 10.18 Experimental coherence spectrum obtained with two anharmonic modulation frequencies $F_1 = 2.0$ and $F_2 = 2.7$. Additional major peaks in the spectrum correspond to "beat frequency" (0.7 Hz), first harmonic of F_1 (4.0 Hz), frequency sum (4.7 Hz), and first harmonic of F_2 (5.4 Hz). The red line represents the same experiment but with 10% of white noise added (Kikas et al., 2001b)

(Fig. 10.19a) and the appearance of breaks shown in Fig. 10.19b. This is an important result because in systems, living or engineered, it may be the only means of the "declaration of success," that is, the actual location of the target (modulation source). It also shows that the range is longer for low frequencies (Fig. 10.19b, lines a and c).
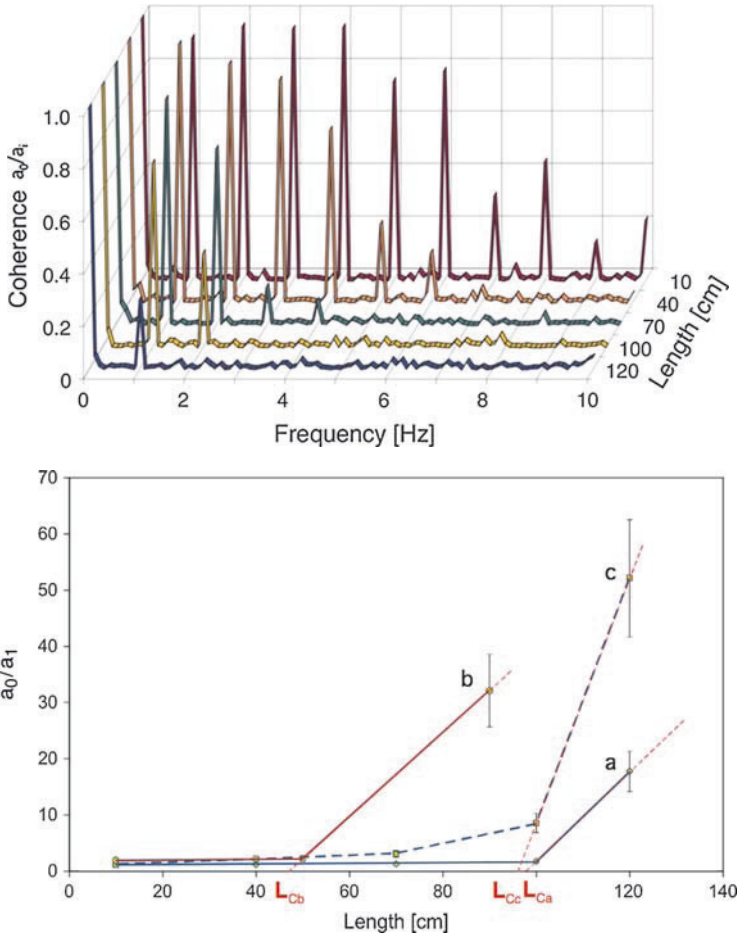
Fig. 10.19 (**a**) Experimental dependence of coherence spectra on distance (10–120 cm) between the modulation source and the sensing array. (**b**) Ratios of intensities a_0/a_1 at fundamental frequency and at its first harmonic, plotted as the function of distance. The breaks indicate "disappearance" of the higher harmonics with distance. They indicate "softening" of plume edges with distance due to diffusion

Similar experiments have also been done with optical sensors, which do not come into physical contact with the plume of fluorescent dye. In this case, the geometrical arrangement of the array and the optimum selection of the appropriate sensor pairs for correlation analysis have played an important role (Fig. 10.20).

As the distance from the modulated source increases, the information contained in the frequencies is washed away and the task of navigation and location of the source must switch to another strategy, one that depends on concentration gradients and amplitudes of the signals. It may be premature to speculate that navigation and signal detection by some living organisms takes place in the frequency, rather than in the amplitude domain (Hartmann, 1998). There is a strong indication that it may
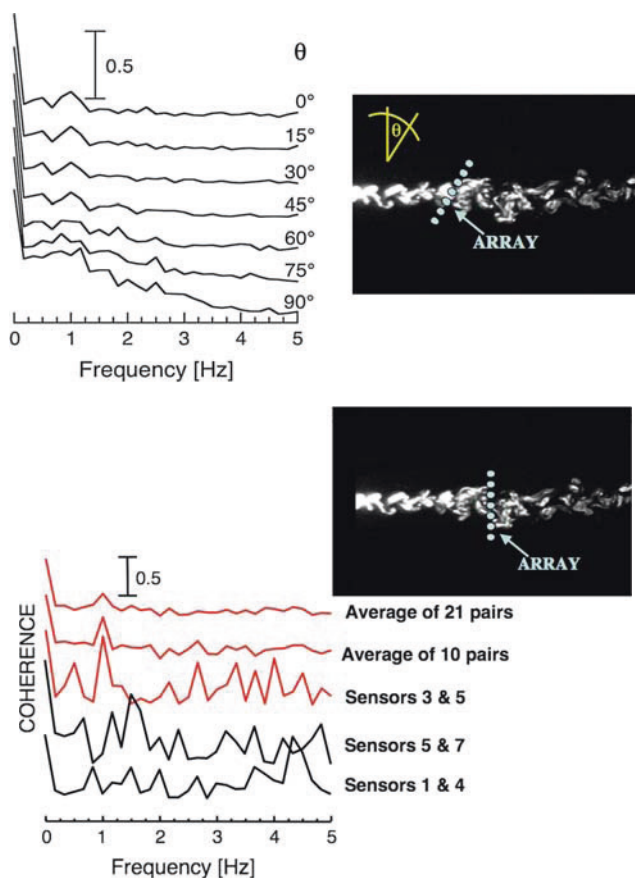
Fig. 10.20 (**a**) Experimental dependence of the coherence spectra on orientation of the linear optical array (Cantor et al., 2008). (**b**) Enhancement of information content achieved by optimized selection of sensor pairs used in the correlation analysis (Cantor et al., 2008)

be so (Weissburg et al., 2002). Nevertheless, there is no doubt that at least at short distances, the information can be encoded/decoded as frequencies and that chemical sensing arrays are capable of detecting it in the dynamic domain.

Food for Thought #10

Beauty Contest

1. There are many—often conflicting—claims of the "best" chemometric method chosen for a given application. Those who make such claims sometimes characterize the competing methods in not very flattering terms. How would you settle the "beauty contest" among various multivariate analysis evaluation methods?

Artificial Neural Networks

2. ANNs have been spectacularly successful in many complex situations. What are the major limitations of their use in the chemical sensing of unknown and possibly variable composition samples?

Symbols

C	Concentration
D	Desired output
d	Distance or diameter
f (or F)	Frequency
m	Number of input elements
N	Number of sensors
n	Number of nodes
$\Re$	Response function
$S(\omega)$	Cross-energy spectral density
t	Retention time
V	Linear velocity
W_n	Weighing factors
X	Sensor response
Y	Actual output
λ	Wavelength
η	Adjustable learning rate
τ	Delay time
γ^2	Coherence

References

Atema, J. (1996) *Biol. Bull*. 191, 129–138.

Cantor, R.S., Ishida, H., and Janata, J. (2008) *Anal. Chem*. 1012–1018.

DiNatale, C., Marco, S., David, F., and D'Amico, A. (1992) *Sens. Act. B* 8, 187–189.

Hartmann, W.M. (1998) *Signals, Sounds and Sensation*. Springer Verlag.

Haykin, S. (1999) *Neural Networks*, 2nd ed. Prentice Hall.

Hierlemann, A., Schweizer-Berberich, M., Weimar, U., Kraus, G., Pfau, A., and Göpel, W. (1996) Pattern recognition and multicomponent analysis. In: Baltes, H., Göpel, W., and Hesse, J. (Eds.) *Sensor Update*, Vol. 2. VCH, Weinheim, pp. 119–180.

Jurs, P.C., Bakken, G.A., and McClelland, H.E. (2000) *Chem. Rev*. 100, 2649–2678.

Kikas, T., Ishida, H., Webster, D.R., and Janata, J. (2001a) *Anal. Chem*. 73, 3662–3668.

Kikas, T., Janata, P., Ishida, H., and Janata, J. (2001b) *Anal. Chem*. 73, 3662–3668.

Kummer, A.M., Burg, T.P., and Hierlemann, A. (2006) *Anal. Chem*. 78, 279–290.

Nakamoto, T., Ishida, H., and Moriizumi, T. (1996) *Sens. Act. B* 35, 32–36.

Osbourn, G.C. and Martinez, R.F. (1995) *Patt. Recogn.* 28, 1793.

Osbourn, G.C., Bartholomew, J.W., Ricco, A.J., and Frye, G.C. (1998) *Acc. Chem. Res.* 31, 207–305.

Ricco, A.J., Crooks, R.M., and Osbourn, G.C. (1998) *Acc. Chem. Res.* 31, 289.

Suslick, K.S. (2004) *MRS Bull.* 720–725.

Weissburg, M.J., Dusenbery, D.B., Ishida, H., Janata, J., Keller, T., Roberts, P.J.W., and Webster, D.R. (2002) *Environ. Fluid Mech.* 2, 65–94.

Appendix A
Survey of Thermodynamics

A.1 Introduction

The laws of thermodynamics are statistical laws. This means that they describe large assemblies of particles called systems. The system is defined as some arbitrary part of the universe with defined boundaries. If neither heat nor matter is exchanged between the system and its surroundings, it is called an *isolated system*. If matter cannot cross the system boundaries, it is said to be a *closed system*; if it can cross, then it is an *open system*. If it is thermally insulated, it is an *adiabatic system*.

Any system can be uniquely described by a minimum number of parameters called state variables. Their relationships are called *equations of state*. For example, an ideal gas inside a piston compartment is a system. It is uniquely described by the equation of state for an ideal gas.

$$pV = nRT \tag{A.1}$$

The state variables are pressure p, volume V, and temperature T. The symbol R represents the universal gas constant and n is the number of moles.

A.2 The First Law

Each system contains a certain amount of internal energy U that resides in the kinetic energy of the individual molecules, in the energy of their bonds, and so on. The First Law of thermodynamics defines the relationship between the work W done by the system ($W < 0$) and the heat Q absorbed by the system ($Q > 0$). Note that the sign of these two variables relates to the exchange of these two quantities with the environment of the system. The First Law simply states that the change of the internal energy of a system dU equals the difference between the work done and the heat received by the system.

$$dU = dQ - dW \tag{A.2}$$

For an adiabatic system, $dQ = 0$ and $dU = -dW$. If the only type of work that a system can do is mechanical work then

$$dU = dQ - P dV \tag{A.3}$$

Another example of work might be electrical work dW_e associated with the moving charge e across a potential difference $d\varphi$.

$$dW_e = e d\varphi \tag{A.4}$$

A third is the work dW_s associated with the increase of surface area dA against surface tension γ

$$dW_s = \gamma dA \tag{A.5}$$

If the chemical reaction is done at constant pressure, the change of internal energy is related to a new thermodynamic function called enthalpy, H.

$$H = U + PV + W_{\text{other}} \tag{A.6}$$

It has a physical meaning of internal energy when the heat is absorbed by the system under constant pressure.

$$dH = dQ_P \tag{A.7}$$

When heat is supplied to the system, it is stored in the kinetic energy of its molecules and the temperature T changes.

$$dT = \frac{dQ_P}{C} \tag{A.8}$$

The actual change of temperature depends on the heat capacity C which is again dependent on whether the process is done under constant pressure (C_P) or at constant volume (C_V). Therefore,

$$C_P = \frac{dQ_P}{dT} = \left(\frac{\partial H}{\partial T} \right)_P \tag{A.9}$$

Both C_P and C_V are constant over only a narrow range of temperatures.

In summary, the First Law of thermodynamics is an empirical accounting statement that relates the internal energy of the system to the heat supplied to it and the work done by it, regardless of the path by which this transaction has been accomplished.

A.3 The Second Law

The Second Law establishes the directionality of the processes taking place in the system by defining a new function, entropy S.

$$dS = \frac{dQ_{\text{rev}}}{T} \tag{A.10}$$

where dQ_{rev} is the reversible change of heat. When a spontaneous (irreversible) process takes place, the entropy of the system increases.

$$\Delta S \geq 0 \tag{A.11}$$

A more intuitive definition of entropy is in terms of probability: a more random system has higher probability and therefore higher entropy.

$$S = k \ln \Omega \tag{A.12}$$

where k is the Boltzmann constant and Ω is the measure of randomness (e.g., probability) of the system. The Gibbs free energy describes all of the useful energy in the system. It is defined as

$$G = H - TS \tag{A.13}$$

For any process to take place spontaneously, the Gibbs free energy must decrease.

$$\Delta G < 0 \tag{A.14}$$

Thus, the aforementioned directionality arises from the combination of (A.11) and (A.13). All closed systems tend to the state of maximum entropy (minimum free energy), that is, the equilibrium.

A.3.1 The Equilibrium

The state of the system at which there is no change of the Gibbs free energy ($dG = 0$) or of the entropy ($dS = 0$), is called the equilibrium. For an isothermal process ($\Delta T = 0$) and two finite states, (A.13) can be restated as

$$\Delta G = \Delta H - T \Delta S \tag{A.15}$$

If $T \Delta S > \Delta H$, the spontaneous process is said to be *entropy-controlled*. The free energy of an ideal solution of species $\mathbf{x}$ is given by

$$G_x = G_x^0 + RT \ln c_x \tag{A.16}$$

Thus, the standard state G^0 is defined for the unit concentration of $\mathbf{x}$. For a simple chemical equilibrium $\mathbf{A} = \mathbf{B}$, the equilibrium constant K is defined as

$$K = \frac{c_B}{c_A} \tag{A.17}$$

It follows from (A.16) and (A.17) that

$$\Delta G^0 = -RT \ln K \tag{A.18}$$

where ΔG^0 is the standard free energy change for the above reaction.

A.3.2 Chemical Potential

The free energy of an open system depends on temperature, pressure, and its composition. If the concentration of component i (number of moles n_i) in the system changes, the free energy of the system changes.

$$\mu_i = \left(\frac{\delta G}{\delta n} \right)_{T,P,j \neq i} = RT \ln a_i \qquad \text{(A.19a)}$$

This partial molar change of free energy is called the chemical potential of species i. If this species is charged (i.e., an ion) it is called the electrochemical potential of species i because the change of the free energy of the system includes a component of electrical work.

$$\widetilde{\mu}_i = \mu_i \pm zF\varphi \qquad \text{(A.19b)}$$

The electrochemical potential of an electron in a (solid) phase $\widetilde{\mu}_i$ is equal to the Fermi potential of that phase. Because at equilibrium, $dG = 0$,

$$dG = \Sigma \mu_i \, dn_i = 0 \qquad \text{(A.20)}$$

This is the Gibbs equation, which is particularly important for understanding phase equilibria. A related expression, called the Gibbs–Duhem equation, states that at equilibrium the change of chemical potential of one component results in the change of the chemical potentials of all other components

$$\Sigma n_i \, d\mu_i = 0 \qquad \text{(A.21)}$$

Concentration of components in a system can be expressed in different units which results in the definition of different standard states (A.16). In a gas phase, it is customary to use partial pressure p_i.

$$p_i = \frac{P_i}{\Sigma P_j} \qquad \text{(A.22)}$$

The most general definition is mole fraction X_i.

$$X_i = \frac{n_i}{\Sigma n_j} \qquad \text{(A.23)}$$

Molarity is the number of moles per one liter of solution and molality is the number of moles per 1,000 g of solvent.

Table A.1 Activity coefficients in solution

$f < 1$	$f > 1$
Electrolytes <0.01 M	Electrolytes >>1 M
Membranes, gels	Organic/water mixtures

A.3.3 Ideal: Real Solutions

Up to this point, we have considered only ideal solutions. If the solute does not behave as an inert particle (i.e., if it interacts with other solutes including its own kind), the energy of this interaction must be included in the internal energy of the system and the (A.16) and those that follow must be defined in terms of activities.

$$a_i = f_i C_i \tag{A.24}$$

The activity coefficient f_i is a measure of the nonideal behavior of species i. For a solution of species i in a noninteracting solvent, it is defined as

$$\lim_{c \to 0} f = 1 \tag{A.25}$$

The activity coefficient can have a value smaller or greater than one.

These interactions exist in real systems for which the activity has the physical meaning of the effective concentration. Thus, only for dilute real solutions does $a_i = c_i$. In mixtures, the activity coefficient is usually, but not always, less than one and is affected by all the species in the multicomponent mixture.

A.3.4 Absorption

When the system contains two (or more) phases the components of the system interact with these phases in one of two ways. The first is that the component exists in both phases, in which case at equilibrium the chemical potential of that component must be equal in both phases, which follows from the Gibbs equation (A.20). Because the number of moles transferred from phase 1 equals the number of moles received by phase $2(\Delta n_1 = -\Delta n_2)$

$$\mu_1 = \mu_2 \tag{A.26}$$

This is the governing law of all partitioning equilibria including the electrochemical, in which the component is absorbed in a phase. In the context of chemical sensors the absorbed species C_{SX} in the selective layer gives rise to the sensor signal. It is related by Henry's law to the concentration of the partitioning species C_S in the sample by partitioning coefficient K.

$$C_{SX} = K C_S \tag{A.27}$$

Equation (A.26) is valid for all components, including the solvent. Because the chemical potential of the solvent is decreased by the presence of solute, the transfer of the solvent across the membrane in the direction of the solvent chemical potential gradient takes place (flow is from the dilute solution to the concentrated solution) with a concurrent increase of the pressure inside the more concentrated solution

compartment. Because pressure is one of the variables ((A.6) and (A.15)) of the free energy, it contributes to the new equilibrium state of the system. This pressure is called *osmotic pressure*.

$$\pi = RTC \tag{A.28}$$

Here C is expressed in moles per liter. It is a colligative property, meaning that it depends on the number of solute molecules. It is also additive, thus the final osmotic pressure is the sum of all partial osmotic pressures.

$$\pi = RT \sum_i C_i \tag{A.29}$$

A.3.5 Adsorption

The second possibility is that the interface between the two phases is impervious to the component (or components), in which case the component adsorbs (accumulates) at that interface. The amount of adsorbed species depends, in a complex way, on the interactions between the adsorbate molecules, between the adsorbate and the interface, and between all the other species competing for the interfacial area. This gives rise to different adsorption isotherms. The simplest is the Langmuir isotherm in which adsorbing species behave as a 2-D ideal gas and do not influence each other (ΔH_{ads} is independent of occupancy). The concentration of adsorbed species C_{SX} is linked by the adsorption equilibrium constant K to concentration C_S of the species in the gas (or solution).

$$C_{SX} = \frac{KC_{SX,max}C_S}{1 + KC_S} \tag{A.30}$$

where the maximum occupation of the surface is described by the term *surface coverage* and the equation $\Theta = C_{SX}/C_{SX,max}$.

$$K = \frac{\Theta}{(1 - \Theta)C_S} \tag{A.31}$$

Interaction between adsorbed molecules and ΔH_{ads} decreases linearly in the Temkin isotherm.

$$C_{SX} = \alpha + \beta \ln C_S \tag{A.32}$$

In the similar Freundlich isotherm, ΔH_{ads} decreases exponentially with occupancy.

$$C_{SX} = aC_S^{1/n} \tag{A.33}$$

In these equations, a, n, α, and β are constants. Adsorption of noninteracting molecules on top of each other to form multiple layers leads to the BET isotherm (Brunnauer, Emmett, and Teller).

$$C_{SX} = \frac{K_1 C_{SX,max_1} C_S}{1 + K_1 C_S} + \frac{K_2 C_{SX,max_2} C_S}{1 + K_2 C_S} + \dots \tag{A.34}$$

where $K_1, K_2, \ldots$ represent the multiple equilibrium constants of consecutive adsorbing layers. In gas the "multilayer" adsorption is a form of "condensation." In that case, ΔH_{ads} decreases with occupancy.

There are many other more or less empirical relationships (isotherms) reflecting different interactions of species in the sample with the selective layer. Combinations of two or more isotherms are also possible. Their general shapes are shown in (Fig. A.1).

A.3.6 Phase Equilibria

A phase is defined as the part of the system that has uniquely distinguishing properties from the other part of the system. That property can be, for example, density (e.g., water–ice–water vapor) or different crystallographic forms (e.g., $\alpha - Pd/\beta - Pd$). The coexistence and number of different phases p depends on the number of components c, and on external physical parameters called degrees of freedom f. These are most typically pressure and temperature. The governing relationship is the Gibbs phase rule.

$$p + f = c + 2 \tag{A.35}$$
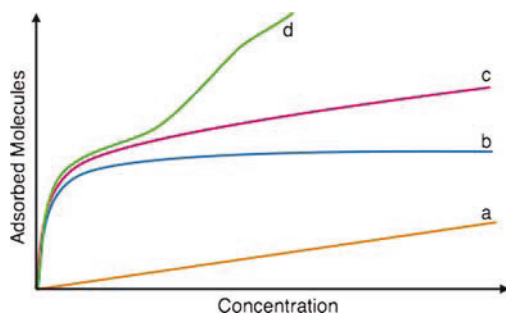
Fig. A.1 General adsorption isotherms
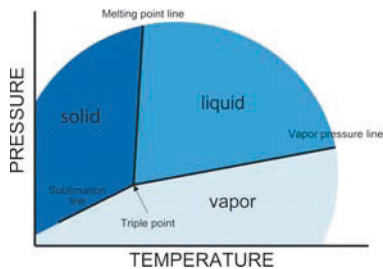
Fig. A.2 Phase diagram indicating the triple point

For example: for a two-phase ($p = 2$) water–ice ($c = 1$) system we have only one degree of freedom ($f = 1$). Thus, we can change the temperature of water and still have coexisting ice, but only at one given pressure. This is the melting point (temperature) A (Fig. A.2) that lies on the coexistence line. If we move to point A' we are in the water phase ($p = 1$) and according to (A.35) we now have two degrees of freedom ($f = 2$), temperature and pressure. This triple point is where all three phases coexist ($p = 3$); it is uniquely defined ($f = 0$). This temperature is the official "zero" of the Celsius scale.

Appendix B
Survey of Kinetics

B.1 Equilibrium and Rate Equations

The *reversible* reaction A ⇔ B can be thought of as being composed of two processes, the forward and reverse. These take place simultaneously with velocity v, thus

$$v_f = -\frac{dC_A}{dt} = k_f C_A \qquad (B.1a)$$

and

$$v_r = -\frac{dC_B}{dt} = k_r C_B \qquad (B.1b)$$

where k_f and k_r are forward and reverse rate constants, respectively.

Because the time change of the activity coefficient is small, the velocity equations are always expressed in concentrations rather than in activities. At equilibrium the two velocities are equal, therefore no net chemical change takes place and (c.f. (A.17))

$$\frac{k_f}{k_r} = \frac{C_B}{C_A} = K \qquad (B.2)$$

Thus, time is introduced as a new variable to the definition of chemical processes including the equilibrium ones. Although the absolute value of K may be the same we can talk about fast or slow equilibria, meaning that the rate constants k_f and k_r are either high or low.

Reactions involving two or more species are reactions of second- and higher-order. Thus for irreversible reaction $A + 2B \rightarrow C$, the rate law is

$$-\frac{dC_A}{dt} = -\frac{1}{2}\frac{dC_B}{dt} = \frac{dC_C}{dt} = kC_A C_B^2 \qquad (B.3)$$

Note that the rate of disappearance of **B** is two times as high as that of species **A** and **C**, hence the $1/2$ coefficient. The reaction is also second-order in **B** because

it depends on the square of concentration of **B**. For the first-order reaction, the dimensions of the rate constant k are s^{-1} whereas for the second-order reaction the dimensions are liters per mole per second. Second- (and higher-) order reactions can be converted to first-order reactions by having all components but one in a large molar excess. Those are called reactions of the pseudo–first-order. For a first-order reaction, the time required to convert exactly one half of the original concentration is called a half-time, $t_{1/2}$, and it is uniquely related to the first-order rate constant.

$$t_{1/2} = \frac{ln2}{k} \tag{B.4}$$

Any reaction scheme can be described by the set of first- or higher-order differential equations. These can be solved exactly for only relatively simple schemes. For more complicated sets, it is necessary to use either analog computers or digital numerical methods.

B.2 Activation Energy

The rate of reaction depends on the temperature and on the value of the activation energy E_a.

$$k = A \exp\left(-\frac{E_a}{RT}\right) \tag{B.5}$$

The value of activation energy determines the magnitude of the rate constant. It is approximately related to the enthalpy of formation of the activation complex, an intermediate in any chemical transformation. Thus, reaction between **A** and **B**, forming **C** can be formally written as progressing through the activated complex **[ABC]***.

$$A + B \rightarrow [ABC]^* \rightarrow C \tag{B.6}$$

For this reaction, the activation energy in (B.5) can be expanded to the corresponding activation entropy ΔS^* and the activation enthalpy ΔH^*.

$$k = A' \exp\left(\frac{\Delta S^*}{RT}\right) \exp\left(-\frac{\Delta H^*}{RT}\right) \tag{B.7}$$

The first exponential term in (B.7) is related to the frequency with which the components of the activation complex assume the correct orientation. It describes, among other things, the stereospecific fit of the components A, B, and C.

The role of a catalyst is to lower the activation energy barrier and hence to increase the reaction rate according to (B.7). On the other hand, inhibitors increase the activation energy and/or cancel the catalytic effect, resulting in lowering of the reaction rate.

B.3 Diffusion

In most sensing situations, the analyte molecules are transported to and from the sensor by diffusion. This is particularly true for sensors in which the analyte is chemically transformed, such as sensors that rely on catalysis or in amperometric sensors. Diffusion is another "activated" (see (B.5)) process in which time and temperature have to be considered. It is driven by the gradient of chemical potential (A.19). Two kinds of diffusion are most relevant. These are the Fickian diffusion, which depends on molecular properties of both the diffusing medium and of the diffusing species, and the Knudsen diffusion which depends on the size of the vessel.

The Fickian diffusion is based on the "random walk" and includes diffusion coefficient D which is related to the dynamic cross-section of the molecule and the frictional losses it experiences as it moves through the medium. The semiempirical Einstein equation (see (7.12)) relates the distance traveled by the molecule in the time interval to the diffusion coefficient.

$$x_i = 2\sqrt{Dt} \tag{B.8}$$

When the molecule diffuses to or from the semi-infinite plane, the rate of diffusion (dC/dt) in direction x is governed by the first Fick's law.

$$\frac{dC}{dt} = -D\frac{dC}{dx} \tag{B.9}$$

The rate of diffusion is also called the flux of the diffusing species. The flux directionality is through an imaginary cross-section which is normal to its direction.

The change of concentration in a volume element of width Δx with time is governed by the second Fick's law.

$$\frac{d^2C}{dt^2} = D\frac{d^2C}{dx^2} \tag{B.10}$$

The solutions of these partial differential equations depend on the boundary and the initial conditions (see, e.g., Section 2.3).

The Knudsen diffusion applies when the mean-free path λ of the molecule is limited by collisions with the wall of a vessel of characteristic physical length L. It is characterized by the Knudsen number K_n.

$$K_n = \frac{\lambda}{L} \tag{B.11}$$

The term is temperature-dependent and contains molecular characteristics of the analyte. In the context of chemical sensing, Knudsen diffusion applies to transport

through porous matrices, gels, and so on, in which the porosity of the matrix is characterized by L.

The diffusion coefficient is related to the radius r of the diffusing molecule and to the viscosity of the medium η by the Stokes–Einstein equation

$$D = \frac{k_B T}{6\pi\eta r} \tag{B.12}$$

Appendix C
Survey of Solid-State Physics

C.1 Introduction

Semiconductors are the base materials of most modern chemical sensors. This appendix surveys the fundamental properties of semiconductors both in isolation and in combination with other materials.

When atoms are "squeezed together" to form a crystal (condensed state), their vacant and filled orbitals become so closely spaced that they merge to form the conduction and the valence energy bands (Fig. C.1). The energy difference between the occupied and filled orbitals is affected by this close packing. The energy difference between the bottom of the conduction band and the top of the valence band is called the bandgap energy E_g.

$$E_g = E_c - E_v \tag{C.1}$$

The magnitude of E_g classifies the materials on the basis of their ability to conduct electricity. For $E_g << kT$ the material is a good conductor and the E_v and E_c overlap. Insulators, on the other hand, have a large value of E_g, typically in excess of 5 eV. Packets of coherent acoustic energy called phonons move through the crystal lattice and cause dissociation of electrons from the valence band to the conduction band leaving behind holes. In an intrinsic semiconductor, the concentration of electrons n (or holes p) is given by

$$n = N_c \exp\left(-\frac{E_g}{2kT}\right) \tag{C.2a}$$

$$p = N_v \exp\left(-\frac{E_g}{2kT}\right) \tag{C.2b}$$

Here N_c and N_v are the effective densities of occupied energy states in the conduction and valence band, respectively. The Fermi level E_F represents the average energy position of electron in the material (Fig. C.2). It is defined as

$$E_F = \frac{E_c + E_v}{2} + 0.5kT \ln \frac{N_v}{N_c} \tag{C.3}$$
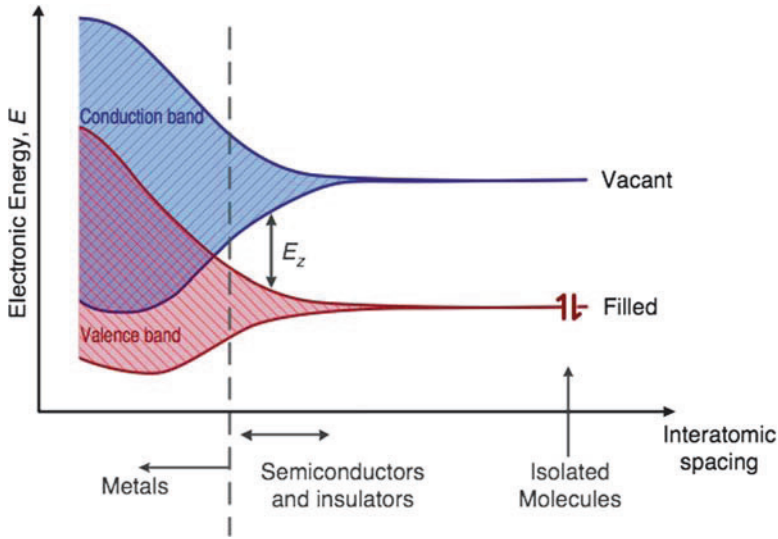
Fig. C.1 Energy bands in various materials

In intrinsic semiconductors $N_\mathrm{v} \sim N_\mathrm{c}$, therefore E_F lies approximately in the middle of the bandgap. For the intrinsic semiconductor, the concentration of holes and electrons is equal because they are generated by the same process. Therefore

$$np = n_i^2 = N_\mathrm{c} N_\mathrm{v} \exp\left(-\frac{E_\mathrm{g}}{kT}\right) \tag{C.4}$$

The analogy between self-dissociation of water to H^+ and OH^- is obvious. Addition of an "impurity" (or dopant) is analogous to the addition of a weak base or a weak acid to water. The same condition of electroneutrality must apply.

$$N_\mathrm{A} + n = N_\mathrm{D} + p \tag{C.5}$$

Here N_A describes negatively charged acceptor ions and N_D describes positively charged donor ions. The excess of holes or electrons then determines the polarity of the semiconductor (Fig. C.2) as p-type or n-type. The position of the Fermi level is affected by the presence of impurities. It lies close to the conduction band for the n-type and close to the valence band for the p-type semiconductor.

$$E_\mathrm{c} - E_\mathrm{F} = kT \ln \frac{N_\mathrm{c}}{N_\mathrm{D}} \tag{C.6a}$$

$$E_\mathrm{F} - E_\mathrm{v} = kT \ln \frac{N_\mathrm{v}}{N_\mathrm{A}} \tag{C.6b}$$

In (C.6a) and (C.6b), N_D and N_A are the concentrations of the donor and acceptor dopants, respectively. The energy required to remove an electron from the Fermi

a

CONDUCTION
BAND

E_C

E_g

E_V

VALENCE
BAND

E_C

E_F

E_V

$n = N_C{}^{-(E_C - E_F)/kT}$
$(= n_i)$

$p = N_V{}^{-(E_F - E_V)/kT}$
$(= n_i)$

b

CONDUCTION
BAND

E_C
E_D

E_V

VALENCE
BAND

E_C
E_F

E_V

n

$np = n_i^2$

p

c

CONDUCTION
BAND

E_C

E_D
E_V

VALENCE
BAND

E_C

E_F
E_V

n

$np = n_i^2$

p

0   0.5   1.0

F (E)

n and p
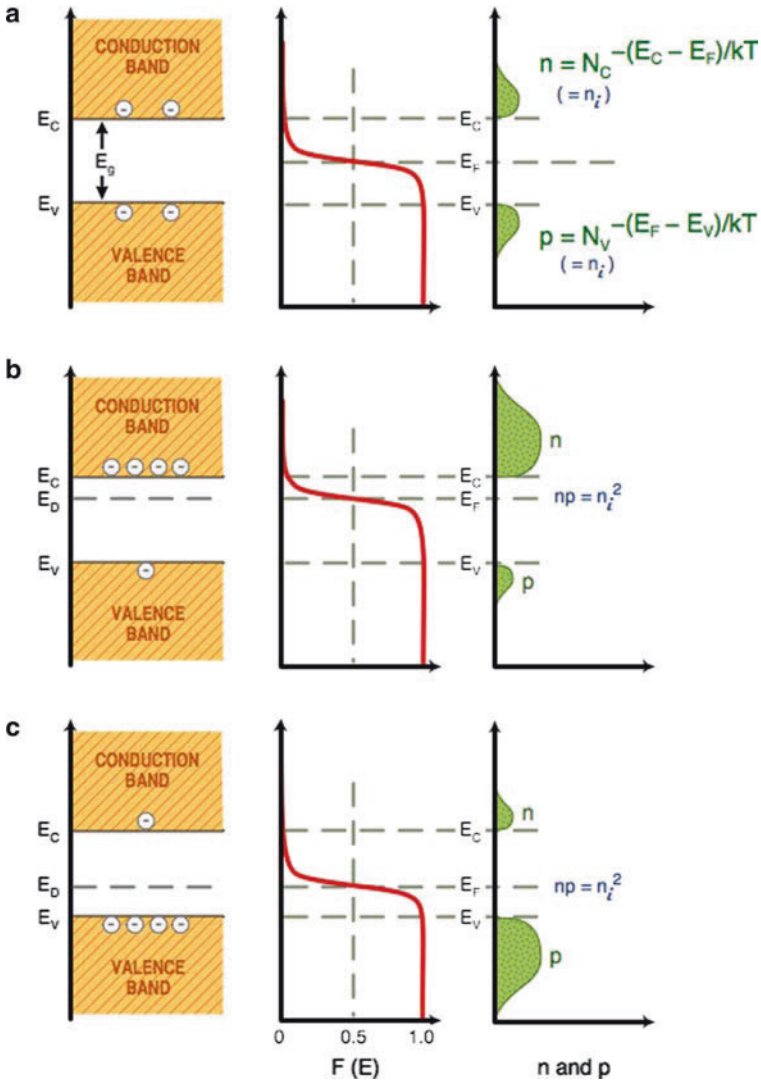
Fig. C.2 Band structure and electron concentration in semiconductors

level and place it in vacuum just outside the reach of the image forces (where $\psi = 0$) is called the work function. It is a material constant defined as

$$\phi = \mu_e - e\varphi = \mu_e - e(\chi + \psi) \tag{C.7}$$

Here φ is the inner electrostatic potential of the bulk, χ is the dipole surface potential, and ψ is the outer electrostatic potential. For $\psi = 0$, the work function consists

of the inner μ_e term and the surface $e\chi$ term. The Fermi level is often called the electrochemical potential of an electron (c.f (A.19), Appendix A).

Charge transport in a semiconductor is done by both the holes and electrons and the material conductivity is defined as

$$\sigma = e\mu_p n_p + e\mu_n n_n \tag{C.8}$$

The mobility of holes μ_p and of electrons μ_n depends on the effective mass m^* of the carriers and on temperature. This is given approximately as

$$\mu = (m^*)^{-3/2} T^{1/2} \tag{C.9}$$

The relationship to the carrier diffusion coefficient D is

$$D_n = \frac{kT}{e}\mu_n \tag{C.10a}$$

$$D_p = \frac{kT}{e}\mu_p \tag{C.10b}$$

In solid-state devices semiconductors are interfaced with other materials and junctions are formed. A junction between two semiconductors of opposite polarity is called a *pn-junction*. If the concentration of one type of dopant is much higher than that of the other (e.g., $N_A >> N_D$) it is called an *abrupt junction* (Fig. C.3). For the approximately equal doping levels we talk about, we use the term *two-sided junction*. A gradient of dopants is found in graded junctions. The concentrations of carriers on the two sides of the junction (subscripted) are

$$p_n = p_p \exp \frac{eV_b}{kT} \tag{C.11}$$

and

$$p_p = p_n \exp - \frac{qV_b}{kT} \tag{C.12a}$$

$$n_p = n_n \exp \frac{-eV_b}{kT} \tag{C.12b}$$

Here V_b is the built-in potential (Fig. C.3). When an external potential is applied, two situations may develop. First, if the polarity of the applied potential follows the polarity of the junction (i.e., positive potential to *p*-type and negative to *n*-type) the junction is said to be forward biased and a current will flow.

$$J_n = -en\mu_n \left(\frac{d\varphi}{dx} \right) \tag{C.13a}$$

and

$$J_p = -en\mu_p \left(\frac{d\varphi}{dx} \right) \tag{C.13b}$$
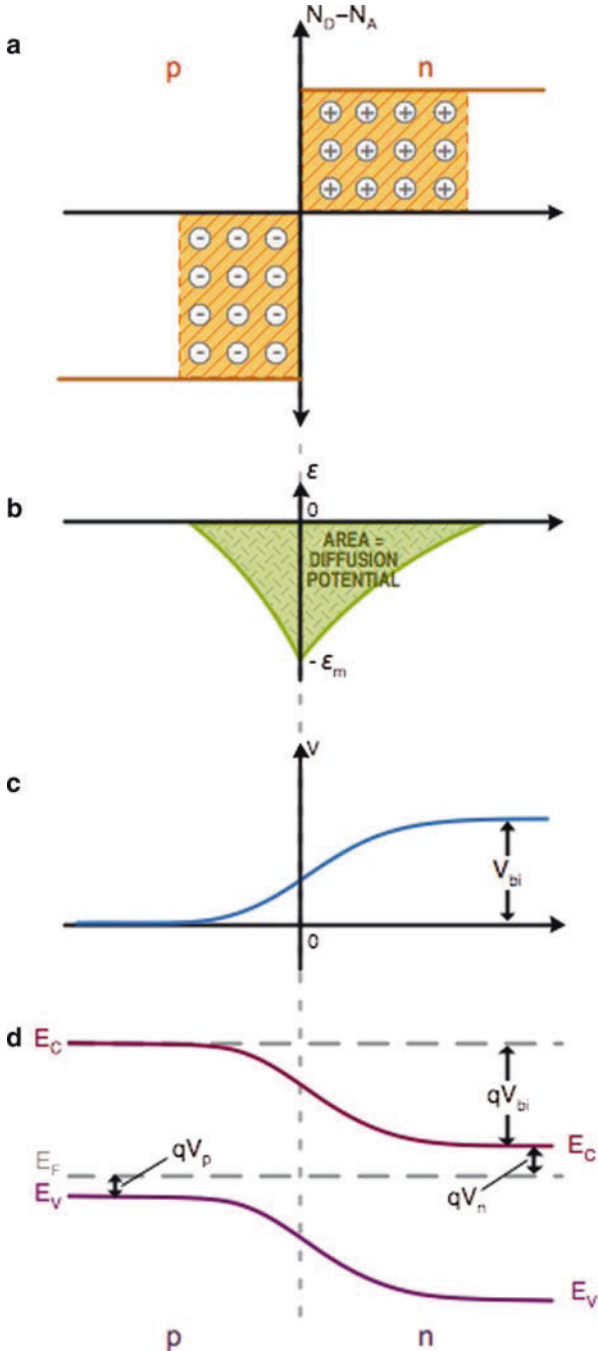
Fig. C.3 Potential junction profiles

For a reverse-biased junction, that is, negative voltage to p and positive to n, a depletion layer will form whose width is given by

$$W = \left(\frac{2\varepsilon_s V_b}{eN_D}\right)^{1/2} \tag{C.14}$$

The capacity of this region is

$$C = \frac{dQ}{dV} = \left(\frac{e\varepsilon_s N_D}{2(V_b - V - kT/e)}\right)^{1/2} \tag{C.15}$$

Rearrangement yields the Mott–Schottky equation. This is used for evaluation of the doping concentration N_D from the plot of inverse square of C versus applied voltage V.

$$\frac{1}{C^2} = \frac{2(V_b - V - kT/e)}{e\varepsilon_s N_D} \tag{C.16}$$

The conditions at the Metal–Insulator–Semiconductor (MIS) junction are discussed in Section C.4 in greater detail. Here we have a look only at the conditions that lead to band bending. This occurs when an external voltage is applied at the MIS junction, creating a residual electric field in the semiconductor. This causes a redistribution of charge and formation of a space charge region near the semiconductor surface. There is a fundamental difference between the space charge caused by the mismatch of the Fermi levels in the two phases (i.e., the Schottky barrier) and the one formed by the passage of current through the junction resulting in formation of a depletion layer. The latter would be described by a diffusion–depletion equation (e.g., (7.16)). They both act as resistances. The surface concentration of electrons and holes is again given by

$$n_s = n_p^0 \exp\frac{e\psi_s}{kT} \tag{C.17a}$$

and

$$p_s = p_p^0 \exp(-\frac{e\psi_s}{kT}) \tag{C.17b}$$

The potential distribution in the space charge region is given approximately by

$$\psi = \psi_s(1 - x/W) \tag{C.18}$$

where ψ_s is the surface potential. The direction of band bending depends on the polarity and magnitude of ψ (Fig. C.4). A special condition exists for $\psi = 0$, when no bending occurs. This is called the flatband condition. It is analogous to the potential of zero charge in electrochemistry.
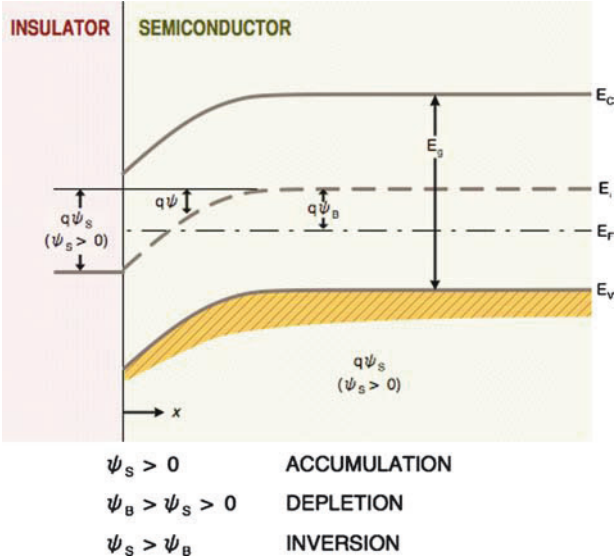
Fig. C.4 Bend bending at the surface of a *p*-type semiconductor. The zero potential is in the bulk of the semiconductor and is referred to the intrinsic Fermi level E_i. The surface potential ψ_s is shown as positive (Sze, 1981)

C.2 The Semiconductor Field Effect

In an extrinsic semiconductor, containing only one type of dopant, there are equal densities of mobile charges (electrons in the *n*-type and holes in the *p*-type) and ionized dopant atoms (positively charged for the *n*-type and negatively charged for the *p*-type). For simplicity, we restrict this discussion to a *p*-type extrinsic semiconductor. The condition of electroneutrality applies only in the absence of an external electric field.

If an electric field is applied to the surface of the semiconductor from whatever source, the density of the mobile charge carriers (holes) is either enhanced or depleted, depending on the polarity of the field. If the field enhances the concentration of holes, the surface is said to be accumulated and the semiconductor surface behaves much as a metal, in that the excess charge appears at the surface and the electric field does not penetrate it further.

If, on the other hand, the field forces the mobile holes away from the surface, a space charge region consisting of the ionized acceptor atoms, which are fixed in the lattice, forms over an appreciable distance into the semiconductor. The thickness of the surface space charge region is a function of the strength of the field at the surface and the semiconductor doping profile, as is the difference between the surface potential and the bulk potential of the semiconductor. If the surface potential deviates sufficiently far from the bulk potential, the surface will invert; that is, it will contain an excess of mobile electrons. In this case, an *n*-type conductive channel

forms near the surface separated from the *p*-type bulk by a space charge region. Further increase of the perpendicular surface field will not significantly change the surface potential, but will only change the density of electrons in the *n*-type channel, and consequently, its electrical conductivity. Changes in the surface electric field can be determined by measuring the corresponding changes in the thickness of the surface depletion region if the surface is not inverted, or the conductivity of the surface inversion layer if the surface is inverted.

The surface field effect can be realized in a number of ways. The semiconductor can be built into a capacitor and an external potential applied (IGFET), or the field can arise from the chemical effects on the gate materials (CHEMFET). In both cases, change in the surface electric field intensity changes the density of mobile charge carriers in the surface inversion layer. The physical effect that is measured is the change in the electric current carried by the surface inversion layer, called the drain current.

C.3 Current–Voltage Relationships for the IGFET

Figure C.5 shows the typical construction and biasing arrangement for the *n*-channel IGFET. In effect, the semiconductor substrate (3) forms one side of a parallel plate capacitor and the metal gate field plate (4) forms the other. In the normal operating mode, the gate voltage V_G is applied between the semiconductor substrate (3) and the gate (4). The polarity and magnitude of V_G are such that the semiconductor field effect gives rise to an inversion layer on the surface of the substrate under the gate. In fact, the surface of the *p*-type substrate (3) becomes *n*-type. This *n*-type inversion layer forms a conducting channel between the source (2) and drain (l) regions. If this inversion layer is not present, application to the drain (l) of a positive voltage V_D with respect to the source (2) and substrate (3) results in no appreciable current flow into drain (l) because the drain-to-substrate $p - n$ junction diode is reverse biased. However, if the *n*-type inversion layer is formed at the surface of the semiconductor, between the source (2) and drain (l), a continuous path for the flow of current exists. The drain current I_D flows through the channel connecting the drain (*l*) and source (2). Its magnitude is determined by the electrical resistance of the surface inversion layer, by the gate voltage V_G, and by the voltage difference V_D between the source and drain.
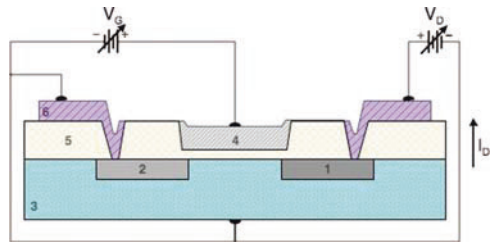
Fig. C.5 Schematic of *n*-channel IGFET: (1) drain; (2) source; (3) substrate; (4) gate plate; (5) field oxide; (6) silicon nitride (Sze, 1981)
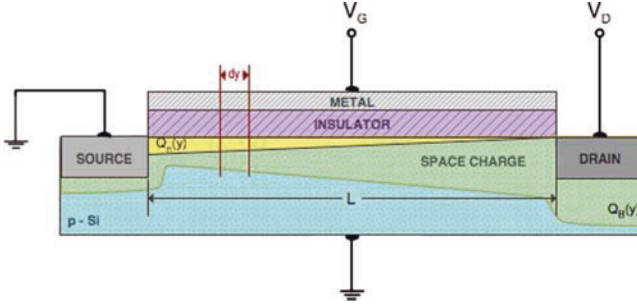
Fig. C.6 Channel region of IGFET (Sze, 1981)

The basis for the derivation of the current–voltage relationships is the calculation of the density of the mobile electrons in the surface inversion layer as a function of the applied voltages V_G and V_D, and position along the channel.

Figure C.6 is a schematic representation of the channel region of the IGFET. At a point y along the channel there is a charge density $Q_n(y)$ per unit area of mobile electrons in the surface inversion layer, and a charge density $Q_b(y)$ of ionized immobile dopant atoms in the space charge region. These charge densities are both functions of y because of the voltage change along the channel due to the drain current. Q_s, the sum of $Q_n(y)$ and $Q_b(y)$, is equal and opposite to the charge per unit area on the metal plate that forms the other side of the capacitor (the gate).

$$Q_s = Q_n(y) + Q_B(y) \tag{C.19}$$

Q_s is related to the applied gate voltage and the capacitance C_0 of the gate insulator by (C.19).

$$V_G - V_{FB} = -\frac{Q_s}{C_0} + \phi_s \tag{C.20}$$

V_{FB} is the flatband voltage defined as

$$V_{FB} = \frac{\phi_M - \phi_{Si} + Q_{ss}}{C_0} \tag{C.21}$$

It accounts for the effects of the difference of work function of the gate metal ϕ_M and silicon ϕ_{Si} and any residual charges Q_{SS} that exist in the gate insulator. ϕ_s is the surface potential of the semiconductor. The mobile electron charge in the surface inversion layer is then given by

$$Q_n(y) = -[V_G - V_{FB} - \phi_s(y)]C_0 - Q_B(y) \tag{C.22}$$

At this point, we assume that the surface potential is given by the condition of strong inversion.

$$\phi_s(y) = V(y) + 2\phi_F \tag{C.23}$$

$V(y)$ is the "reverse bias" applied in the perpendicular direction to the "field-effect induced" p–n junction created by the n-type surface inversion layer and the p-type substrate. The charge $Q_B(y)$ created by the immobile ionized atoms of the dopant forms the space charge region. The "depletion approximation" assumes that there are no mobile charges at all in the space charge regions. Under that assumption, (C.24) applies.

$$Q_B(y) = -\sqrt{2\varepsilon_{Si}\varepsilon_0 eN_A[V(y) + 2\varphi_F]} \tag{C.24}$$

It is seen that $Q_B(y)$ increases with $V(y)$, the voltage drop along the channel. This increases at the expense of $Q_n(y)$, the mobile channel charge. For the sake of simplicity, it is possible to neglect the weak dependence of $Q_B(y)$ on the voltage drop (C.22), which only marginally affects the accuracy of these equations. In that case

$$Q_n(y) = -\sqrt{V_G - V_{FB} - \phi_s(y)]C_0 + 2(\varepsilon_{Si}\varepsilon_0 eN_A \phi_F} \tag{C.25}$$

Referring to Fig. C.7, the voltage drop along the element of channel, dy, can be written as

$$dV = I_D dR = I_D \frac{dy}{W\mu_n Q_n(y)} \tag{C.26}$$

W is the width of the channel and μ_n is the electron mobility in the channel. Substitution of the above expression for $Q_n(y)$ and integrating over the channel length yields the basic current voltage relationship.

$$I_D = \frac{\mu_n C_0 W}{L} V_D(V_G - V_T - V_D/2) \tag{C.27}$$

where the threshold voltage V_T is

$$V_T = V_{FB} + 2\phi_F - Q_B/C_0 \tag{C.28}$$
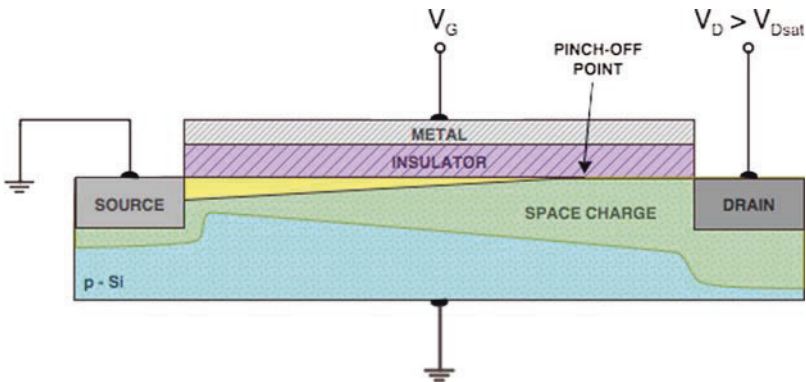
Fig. C.7 IGFET operated in saturation mode (Sze, 1981)

In the derivation of (C.26), it was assumed that the surface inversion layer actually existed at all points along the channel. Because $V(y)$ must change from 0 at the source to V_D, the conducting channel disappears near the drain end of the channel if V_D is large enough. When this happens, (C.26) is no longer valid and the IGFET is said to become saturated. Figure C.7 is a schematic diagram showing the IGFET in saturation mode. V_{DSAT} is the drain voltage at which the surface channel just disappears at the drain end.

$$V_{DSAT} = V_G - V_T \tag{C.29}$$

Any increase in V_D beyond V_{DSAT} results in the formation of a short space charge region between the channel and the drain. Under these conditions, drain current continues to flow because electrons in the channel do not see any potential barrier restricting their flow from the channel across the space charge region to the drain. In other words, the transit of electrons across the drain electrode space charge becomes ballistic. However, the number of electrons arriving at the drain end of the channel is determined by the voltage V_{DSAT}, between the source and the end of the surface channel. For drain voltages greater than V_{DSAT}, the drain current I_D is given by (C.30) with I_D replaced by V_{DSAT}, providing the channel length L is not appreciably shortened by the space charge region between the end of the channel and the drain. This is called the long channel approximation.

$$I_D = \frac{\mu_n C_0 W}{2L}(V_G - V_T)^2 \tag{C.30}$$

The I_D–V_D curves in Fig. C.8 are divided into the saturated and unsaturated regions. The threshold voltage is typically evaluated from the linear form of (C.30).

$$\sqrt{I_D} = \left(\frac{\mu_n C_0 W}{2L}\right)^{1/2}(V_G - V_T) \tag{C.31}$$
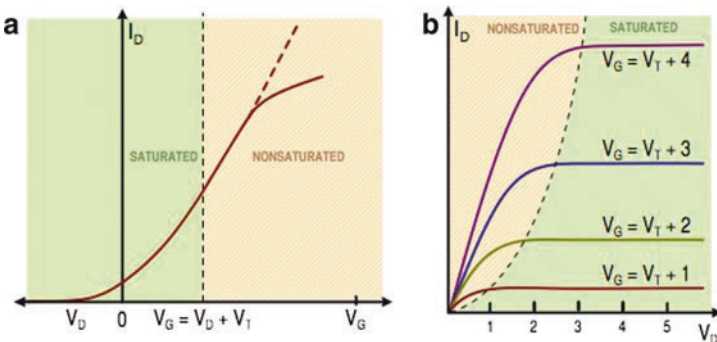
Fig. C.8 Characteristics of an *n*-channel IGFET: (**a**) plot of drain current versus gate voltage; (**b**) plot of drain current against drain voltage for various applied gate voltages (Sze, 1981)

Equations (C.27) through (C.31) have the distinct advantage of simple form and are the IGFET current–voltage relations most often quoted in the literature. Their use can lead to significant quantitative errors in calculated currents. They do, however, contain the correct qualitative features of the device: the saturated and unsaturated regions, the near constant drain current beyond saturation, and the dependence of the saturation voltage on the gate and threshold voltages.

Figure C.8A is a plot of I_D versus V_G for an n-channel depletion mode device. Depletion mode refers to the fact that the device has a conducting channel for zero applied gate voltage. It is the result of a sufficiently negative flatband voltage, V_{FB}. In the saturated regions, that is, $V_D > V_G - V_T$, or alternately $V_G < V_D + V_T$, (C.30) predicts a quadratic dependence of I_D on V_G. In the unsaturated region, that is, $V_D < V_G - V_T$ or alternately $V_G > V_D + V_T$, (C.27) predicts a linear dependence of I_D on V_G. Actual data, however, show a significant departure from linearity in the unsaturated region as indicated in Fig. C.8A. One cause of this departure from theory is the variation of μ_n, the charge carrier's mobility, with the electric field normal to the surface of the semiconductor. Surface channel mobility is primarily determined by surface scattering. Strong electric fields that increase the probability of the carrier interacting with the surface reduce the surface mobility. Therefore, as the gate voltage increases, the electric field normal to the surface increases and μ_n decreases. This condition is called velocity saturation. A second cause of the departure of the $I_D - V_G$ curve from linearity is series resistance between the end of the channel and the point at which the drain voltage is actually measured. The resistive voltage drop actually reduces the effective drain voltage below that applied to the device. The magnitude of this resistance depends on both the individual device geometry and processing parameters, but it may be quite significant, up to $200\,\Omega$ in some designs.

As a result of the effects of nonideal structures, second-order effects in parameters, and the numerous approximations made in the derivation of the current–voltage equations, (C.27) and (C.30) can only serve as a qualitative description of the actual device; each individual design must be experimentally characterized. For these reasons it is advantageous to operate the FET in the constant drain current mode in which case a suitable feedback circuit supplies the gate voltage of the same magnitude but of the opposite polarity to that produced by the electrochemical part of the device.

The fact that IGFET can be perceived as a field-dependent resistor is seen from the expression for transconductance g_m which is defined as the average conductance of the channel for a given V_D. For nonsaturation from (C.27)

$$g_m = \left(\frac{dI_D}{dV_G}\right)_{V_D} = \frac{W\mu_n C_0 V_D}{L} \tag{C.32}$$

and from (C.30) for saturation

$$g_m = \left(\frac{dI_D}{dV_G}\right)_{V_D} = \frac{W\mu_n C_0 (V_G - V_T)}{L} \tag{C.33}$$

It is useful to summarize the assumptions made in the derivation of (C.27) and (C.30).

(a) It is assumed that there exists a well-defined threshold voltage, and that the formation of the surface inversion layer begins suddenly as the gate voltage is increased. It is equivalent to stating that there is a sharply defined semiconductor surface potential dividing surface depletion and surface inversions. In fact, this transition is continuous. In conventional structures this is a good approximation if the gate voltage exceeds the "threshold voltage" by about 0.56 V.

(b) The voltage dropped due to a surface channel current (drain current) flowing has no effect on the thickness of the surface space charge region. This approximation can lead to relatively large errors in the magnitude of the predicted drain current in the saturation region, but the general shape of the drain current versus drain voltage curves is satisfactory; that is, the qualitative features of the device are not affected by this approximation.

(c) The doping of the semiconductor is constant near the surface where the channel is formed. This is not a good approximation for many ion-implanted structures.

(d) The channel length is large compared to the thickness of the depletion region surrounding the $p-n$ junctions.

(e) Both the source and substrate of the device are connected to the same point of the external circuit.

(f) The thickness of the space charge region along the channel is constant.

Reference

Sze, S.M. (1981) *Physics of Semiconductor Devices*. Wiley.

Appendix D
Conversions of Equivalent Electrical Circuits

Commercial impedance analyzers offer equivalent circuit interpretation software that greatly simplifies the interpretation of results. In this Appendix we show two simple steps that were encountered in Chapters 3 and 4 and that illustrate the approach to the solution of equivalent electrical circuits. First is the conversion of parallel to series resistor/capacitor combination (Fig. D.1). This is a very useful procedure that can be used to simplify complex RC networks. Second is the step for separation of real and imaginary parts of the complex equations.

It may be counterintuitive that a resistor and a capacitor in series combination have different nominal values than a resistor and a capacitor connected in parallel. What we are seeking is the equivalency of the two combinations; that is,

$$Z_s = Z_p \tag{D.1}$$

The impedance of the series combination Z_s is a complex number.

$$Z_s = R_s + 1/j\omega C_s \tag{D.2}$$

Here ω is the angular frequency and $j = \sqrt{-1}$. For the parallel combination we write

$$1/Z_p = 1/R_p + j\omega C_p \tag{D.3}$$

Let us define for series s and parallel p combination auxiliary parameters W_s and W_p, respectively,

$$W_s = (\omega C_s R_s)^2 \tag{D.4a}$$

and

$$W_p = (\omega C_p R_p)^2 \tag{D.4b}$$

The admittance of the series combination from Eq. (D.2) is

$$Y_s = \frac{j\omega C_s}{j\omega R_s C_s + 1} \tag{D.5}$$
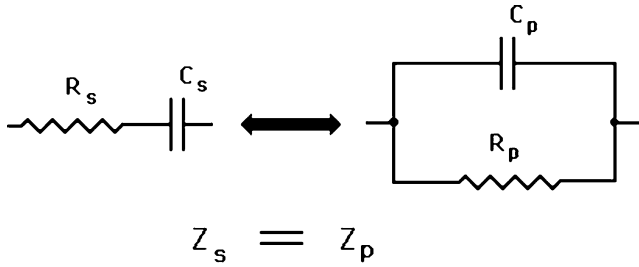
Fig. D.1 Conversion of series to parallel RC combination

When both denominator and numerator are multiplied by $((j\omega R_s C_s) - 1)$, the real and the imaginary components of (D.5) separate, and we get

$$1/Z = \frac{\omega^2 C_s^2 R_s}{(\omega C_s R_s)^2 + 1} + \frac{j\omega C_s}{(\omega C_s R_s)^2 + 1} \tag{D.6}$$

This separation step is based on the relationship $(x + y)(x - y) = (x^2 - y^2)$ and is used when denominator is a complex number. Combining (D.4) and (D.6) yields

$$1/Z = \frac{W_s/R_s}{W_s + 1} + \frac{j\omega C_s}{W_s + 1} \tag{D.7}$$

For the two combinations to be equal, the real and the imaginary components of their impedances must be equal for all frequencies. Thus, from (D.3) and (D.7) for the real component

$$1/R_p = \frac{W_s/R_s}{W_s + 1} \tag{D.8a}$$

or

$$R_p = R_s \frac{W_s + 1}{W_s} \tag{D.8b}$$

and for the imaginary component

$$C_p = \frac{C_s}{W_s + 1} \tag{D.9}$$

A similar analysis can be performed for the series combination, starting from the parallel combination. The resulting formulas are as follows.

$$R_s = \frac{R_p}{W_p + 1} \tag{D.10}$$

$$C_s = \frac{C_p(W_p + 1)}{W_p} \tag{D.11}$$

Index