

357B

745685

2065725

TR3571

Characterization of coatings inside food and beverage cans

Ardjan de Jong

Faculty of Applied Sciences
Delft University of Technology
October 2000

Cover design: Ardjan de Jong, Cécile Combes and Bart Norbart
Several tools designed and used for this research project

Characterization of coatings inside food and beverage cans

Proefschrift



ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.F. Wakker,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen

op maandag 2 oktober 2000 om 10:30 uur

door

Adriaan Jan DE JONG

natuurkundig ingenieur
geboren te Delft

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. J.H.W. de Wit

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter	
Prof. dr. J.H.W. de Wit,	Technische Universiteit Delft, promotor
Prof. dr. J.D. Fehribach,	Worcester Polytechnic Institute, USA-Worcester
Prof. dr. ir. J. van Turnhout,	Technische Universiteit Delft
Prof. dr. G. den Ouden,	Technische Universiteit Delft
Prof. dr.-ing. H.D. Liess,	Universität der Bundeswehr München, D-Neubiberg
Dr. K. Hemmes,	Technische Universiteit Delft
Dr. J.H.O.J. Wijenberg,	Corus, IJmuiden

This research project was carried out in co-operation with and sponsored by Corus Research, Development & Technology, Centre for Packaging Technology.

Published and distributed by:

Delft University Press
P.O. Box 98
2600 MG Delft
The Netherlands
Telephone: +31 15 2785121
Telefax: +31 15 2781661
E-mail: DUP@Library.TUdelft.NL

ISBN 90-407-2097-5

Copyright © 2000 by A.J. de Jong

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the publisher: Delft University Press.

Printed in The Netherlands

Contents

1. Introduction	1
1.1 Short history of food and beverage cans	1
1.2 The prefinished approach of can manufacturing	2
1.3 Corrosion in food and beverage cans	3
1.4 Testing the corrosion protection capabilities of polymer coatings.....	4
2. Diffusion of water and ions through polymers and some measurements on PET	7
2.1 Introduction	7
2.2 Diffusion in polymers.....	7
2.2.1 Fick's first law	8
2.2.2 Diffusion of water and ions in polymers.....	12
2.2.3 The diffusion equation, Fick's second law	14
2.2.4 Diffusion of water and ions in polymer films.....	15
2.3 Diffusion measurements.....	17
2.3.1 Gravimetric measurements	19
2.3.2 Electric impedance measurements.....	21
2.4 Water and ion diffusion measurements	28
2.4.1 Results on water and ion diffusion in PET from literature	29
2.4.2 Gravimetric diffusion measurements on PET films.....	30
2.4.3 Electric impedance measurements of water and ion uptake	32
2.4.4 Long-term measurements on water and ion transport in PET	39
2.5 Summary and discussion.....	41
3. EIS on artificial defects in PET coated ECCS	43
3.1 Introduction	43
3.2 Electric impedance spectroscopy	43
3.2.1 Systems and signals	44
3.2.2 Real signals and systems.....	47
3.2.3 Electric impedance spectroscopy.....	50
3.3 Measuring the adhesion of coatings on metals by EIS.....	56
3.3.1 Delamination of polymer coatings from metals in electrolytes.....	56
3.3.2 Characterization of coated metals in electrolytes	57
3.3.3 Characterizing the delamination of coatings.....	61
3.4 The delamination of PET coated ECCS in citric acid	63
3.4.1 Creating artificial defects.....	63
3.4.2 Iron in citric acid.....	66
3.4.3 Delamination measurements on PET coated ECCS	68
3.5 Conclusions and suggestions for future research	73

4. Deformation of PET coated ECCS during a DWI-process	75
4.1 Introduction	75
4.2 Geometrical model for a DWI-process.....	75
4.3 Some measurements on cans during a DWI-process	78
4.3.1 Height of the defects in the cups.....	79
4.3.2 Deformation of the cups.....	80
4.4 Summary and discussion	84
5. The application of electric potential scans for studying localized corrosion	87
5.1 Introduction	87
5.2 The electric field above a corroding surface	88
5.2.1 Metals	88
5.2.2 Electrolytes on metals	91
5.2.3 The electric potential in the bulk around interfaces.....	99
5.2.4 Coatings	100
5.3 Summary and conclusions.....	101
Appendix 5A. The Nernst equation.....	102
Appendix 5B. The Butler-Volmer equation.....	104
6. Measuring with the Scanning Kelvin Probe	109
6.1 Introduction	109
6.2 The Kelvin Probe method.....	110
6.2.1 Homogeneous surfaces	110
6.2.2 Inhomogeneous surfaces.....	113
6.2.2.1 Data enhancement by inversion techniques	115
6.2.2.2 Stray capacities.....	117
6.2.3 Distance control	117
6.2.3.1 Distance control with the capacity	118
6.2.3.2 High U distance control.....	118
6.2.3.3 Second harmonic distance control.....	119
6.2.3.4 Slope distance control	119
6.2.3.5 Other methods of distance control.....	120
6.2.4 The influence of insulating and poorly conducting coatings	122
6.2.4.1 Perfectly insulating coatings.....	122
6.2.4.2 The influence of an insulating coating on distance control.....	124
6.2.4.3 The influence of the coating conductivity.....	126
6.3 Surface potential measurements.....	127
6.3.1 Standard tests on metal surfaces	127
6.3.1.1 Stability of the measured potential.....	127
6.3.1.2 Measurements of a well-defined voltage.....	127
6.3.1.3 Measurements of the resolution	128
6.3.1.4 The influence of stray capacities	129
6.3.1.5 Distance control.....	130
6.3.2 Metal-electrolyte systems	132
6.3.2.1 Copper in a CuSO_4 solution in water	133
6.3.2.2 Iron in diluted sulfuric acid	133
6.3.3 Coated substrates	134

6.3.3.1	PET coated ECCS	134
6.3.3.2	Glass on metals.....	138
6.3.3.3	Scotch tape® on metals.....	139
6.3.3.4	Underpaint corrosion.....	141
6.4	Summary and conclusions.....	142
7.	Inversion of dielectric spectra for obtaining time constant distributions	145
7.1	Introduction	145
7.2	Three inversion methods	147
7.2.1	Least square fitting.....	147
7.2.2	Inversion by differential operators.....	148
7.2.3	Deconvolution by Fourier transforms.....	150
7.3	Limitations of the inversion algorithms	154
7.4	Comparison of the performance of the inversion algorithms.....	156
7.4.1	Cole-Cole data without noise.....	156
7.4.2	Spectra with noise.....	157
7.5	Summary and discussion	158
	Appendix 7A. A factorized differential operator.....	171
	References	177
	Summary	183
	Samenvatting	187
	List of Publications	191
	Curriculum Vitae	193
	Dankwoord	195



Chapter 1

Introduction

1.1 Short history of food and beverage cans

In 1791, the Frenchman Nicolas Appert found that food could be preserved by sealing it in an airtight glass jar and then heating it up to 100°C. He reached his conclusions after a period of 40 years of research. In 1804, he started a factory to explore his method and already around 1806, the French navy was carrying his jars, filled with meat, vegetables, fruit and milk. In 1809, Appert published his results and, according to the legend, he received an award of 12,000 Francs from the government of Napoleon Bonaparte.

But although the heating method of Appert clearly worked, it was not understood why. This was only discovered in the second half of the 19th century by Louis Pasteur, who first studied the role of microorganisms in the decay of food and beverages. For instance, he found in 1864, on request of Napoleon III, that a wine disease was caused by microorganisms and could be overcome by heating the wine to 55°C for several minutes. Nowadays, such heating processes with a temperature below 100°C are called “pasteurization”.

In England, Peter Durand patented the conservation of food and beverages in containers of all kinds. He sold his rights to the Donkin, Hall & Gamble company, who introduced tins for food conservation in 1813. At that time, tins were produced by dipping steel sheet in molten tin. In the 20th century, the development of tins and its uses proceeded, including the introduction of the continuous electroplating process of steel strip. Tins were much less breakable than glass jars, but were rather costly. They were made by hand from three pieces of tinplate (the bottom, the top and a cylindrical bend sheet forming the wall) and even a skilled person could not make more than 6 cans per hour. The situation improved when towards the end of the 19th century, the first can making machines were developed. Nowadays, one beverage can-making line produces about 2000 cans per minute.

In the 1950s, tin became scarce and in search for other can materials, chromium plated steel, or ECCS (Electrolytic Chromium Coated Steel) was introduced (cf. [Hartman 1992]). Since ECCS corrodes in contact with most food and beverages, it is protected by a polymer coating, such as an epoxy lacquer. The upper chromium-oxide layer of ECCS

provides good adhesion to the polymer coating. Another alternative can material was aluminum, introduced in 1958 by the Adolf Coors Company. An additional innovation was that the aluminum cans were made from two pieces instead of three, which meant a considerable simplification of the can making process (cf. [Hosford 1994]). The two-piece can basically consists of a can body, which is made from a plane sheet by Drawing and Wall-Ironing (the so-called “DWI process”) and a top. In 1974, the two-piece can concept was also introduced for tinsplate cans (cf. [Newman 1999]). Due to strong restrictions on the metal content in food and beverages, one started to protect cans with polymer coatings in the 1970s. A number of different coatings have been used for that purpose (cf. [Montenari 1996]).

Aluminum and tinsplate have both advantages and disadvantages for the application as food and beverage containers. The most important considerations are that under natural conditions, aluminum has a better corrosion resistance, while the production of steel by cokes reduction is energetically more economical than the electrolysis of bauxite by which aluminum is produced (cf. [Hosford 1994]). In the US, aluminum cans dominate the market while in Europe, the situation between tinsplate and aluminum cans is about fifty-fifty [Hill 1996]. The top of beverage cans is always aluminum, as required for the “easy opening system”, introduced in 1960. Recently, easy opening systems for steel cans have also been developed (cf. [Hartman 1992]). Interestingly enough, aluminum is basically used for beverage cans. Most food cans are still three-piece cans of tinsplate, but produced by modern high-speed welding equipment (cf. [Sodiek 1988]).

1.2 The prefinished approach of can manufacturing

Currently, two-piece beverage cans are made from metal plate and coated afterwards. This requires a proper cleaning of the can after the DWI-process, and one or two sprayings and bakings with the epoxy coating. A new development is the “prefinished approach” (cf. [Aa, van der, 1999]), in which case the metal sheet is coated before forming the can. This not only means a significant simplification of the can manufacturing process, but also prevents the use of organic solvents in case so-called polymer resins such as polyesters are used.

The prefinished approach demands is very demanding for the coating strength and it is almost unbelievable that polymer coatings may withstand the DWI-process. Nevertheless, it was found at Corus RD&T that in combination with ECCS, Polyethylene Terephthalate (PET) is promising for the prefinished approach.

PET is a well-known material in the packaging industry, especially since the introduction of the PET-bottles for beverages. It is one of the many new polymers that were discovered during the Second World War. Under the trade name Mylar, DuPont commercially produced it around 1950. At that time, PET was basically used for fibers. In fact, nowadays still 75% of the entire PET production is directed towards fiber manufacturing and used PET bottles are recycled as fibers to make winter coats and jumpers (the so-called “fleece-jumpers”). There are several reasons why PET is very well suited for the storage of food and beverages. It has no taste, it is relatively strong and being a thermoplastic polymer, it can be easily formed into the shape of a bottle (by blowing it into a mould at elevated temperature). In addition, PET only contains the elements C, H and O and is therefore an environmentally friendly polymer.

One part of the research on PET coated ECCS for the prefinished approach is the optimization of the composition of the PET layer and the parameters of the DWI-process, such that the coating is not damaged and keeps its protective properties [Aa, van der 1999]. Another part of the work is to study the performance of the produced cans for the actual conservation of food and beverages.

1.3 Corrosion in food and beverage cans

The conservation of food or beverages means that their quality is maintained for a specified period, which is called the “shelf life”. In general, the quality of a product can decay:

- Chemically (mostly by oxidation of the product);
- Physically (such as by drying);
- Microbiologically (modification of the product by the action of bacteria or fungi);
- Pickup of ions from the can metal.

Steel and aluminum are impermeable to oxygen and water, so the first two problems are a matter of properly closing the can, while destroying microorganisms is a matter of an effective heat treatment. The pickup of metal ions is due to corrosion and is the subject of interest of this work.

Corrosion inside a can is often influenced by the stored food and beverage products themselves, which are generally rather corrosive due to a low pH (typically between 2.5 and 6) and because they are normally good electrolytes, since they contain a large amount of dissolved ions. On the other hand, the corrosion processes in cans are slowed down by the low oxygen contents of the electrolyte. The oxygen concentration of the electrolyte is low, because the oxygen is largely removed when the can is filled. This is primarily done to prevent oxidation of the stored product, but as a side benefit it also slows down the corrosion of the can metal at these pH values (cf. [Uhlig 1963]).

Another important issue for can corrosion is the initial heat treatment such as pasteurization or sterilization. This is done when the product is canned and is therefore undergone by the can material too. Coating adhesion should be maintained during such a treatment. Not all food and beverages receive an initial heat treatment. As a rule of thumb, this is not necessary for food and beverages with a pH below 3.5, such as for most carbonated soft drinks. There are however exceptions. For instance, some producers pasteurize beer, while others do not.

For the prefinished approach finally, the influence of the can formation process on the corrosion resistance of the coating is relevant. Summarizing, we have to consider the following issues for studying the corrosion behavior of PET coated ECCS:

- The low pH of the product;
- The high electric conductivity of the product;
- The low oxygen content of the product;
- The initial heat treatment of the canned product;
- The deformation of the coating.

One could also mention the influence of the pressure inside the can, which can be up to 6 bar in case of carbonated beverages. At these high pressures, the amount of dissolved gases in the electrolyte can be higher. For corrosion however, oxygen is the most relevant gas and since most of the oxygen is initially removed, the pressure will generally not have much influence on the corrosion processes inside a can.

For doing well-defined experiments, one often prefers to work with model electrolytes instead of real food or beverage products. Buffered citric acid with $\text{pH} = 3$ is a standard model electrolyte, which corresponds to soft-drinks based on citric acid, such as Fanta. Another common model electrolyte is phosphoric acid that models a Cola drink.

1.4 Testing the corrosion protection capabilities of polymer coatings

A reliable test for the corrosion protection afforded by a coating inside a can is to store the food or beverage product of interest in a large number of cans and to check the metal contents in the product after an exposure time equal to the shelf life. The metal uptake of the product can be determined by Atomic Absorption Spectrometry (AAS) (cf. [Bird 1971, Kleniewski 1975, Dolfig 1995, Seruga 1996]). However, this test is time consuming and not very practical in the development stage where one would like to rapidly examine a large number of different coatings. Therefore, considerable research has been done to methods of predicting the metal uptake of food and beverage products by determining the corrosion rate in a much shorter time. Electrochemical Impedance Spectroscopy (EIS) has been especially useful for this purpose (cf. [Bird 1971, Kleniewski 1975, Grandle 1994, Bovard 1995, Dolfig 1995, Bugnard 1996, Galic 1998, Pezzani 1998]).

Another limitation of this shelf life exposure test is that it does not give information on the protective mechanism of the coating. Such information is of course also interesting for the development of coatings. In general, the corrosion protection of polymer coatings is based on the fact that they form a barrier between the metal and the environment. Thus they prevent the metal to come in contact with corrosive electrolytes. The barrier functions are in fact:

- Slowing down the transport of the electrolyte to the metal surface. Here, the transport of water, ions and gases must be considered separately. For corrosion in cans, the transport of water and H^+ is most significant, since most of the oxygen is initially removed;
- A good adhesion to the metal surface, such that no metal-electrolyte interface can be formed, even when the coating is finally saturated with electrolyte (wet adhesion).

Which of the two is most important depends on the type of coating. In some cases, the ion transport is found to be rate determining (cf. [Bacon 1948, Mayne 1970, Cherry 1974]) while in other cases, the adhesive properties of the coating are found more important (cf. [Walker 1967, Wit, de 1995]). Most authors however agree that in general, the transport of electrically neutral molecules such as water and oxygen is not rate determining for corrosion. This is also illustrated by permeability measurements on free polymer films. (For a table of permeability parameters of different polymer coatings, cf. [Wit, de 1995].)

In order to measure the adhesion of a coating, the standard Deutsche Industrielle Norm test (generally abbreviated by "DIN-test") is to apply a piece of adhesive tape on the

coating and then to rapidly remove the tape. Also for the electrolyte transport through the coating, DIN-tests exist, such as penetrant-testing by a CuSO_4 -solution in water. In the latter case, the copper ions precipitates on places where the CuSO_4 -solution can reach the underlying metal and become visible as red spots (For a more detailed description of the relevant DIN-tests, cf. [Montenari 1996]). A limitation of these DIN-tests is that they do not give quantitative results on the adhesion of the coating and that they can only be used for detecting macroscopic defects in coatings and not for the characteristic micropores in the whole polymer.

One of the objectives of the work of this thesis was to study the electrolyte transport through polymers and the adhesion of polymers on metals by more sophisticated methods. In chapter 2, a study on the electrolyte transport in PET is presented by measuring the dielectric properties of PET after immersion by electrolyte. The adhesion of PET on ECCS is studied in chapter 3 by following the growth of well-defined artificial defects. These were round holes in the PET coating with a diameter of about $60 \mu\text{m}$. The methods of chapters 2 and 3 are not standardized in industry and therefore, considerable attention has been paid to the implementation and interpretation of the methods themselves. Artificial defects in PET on ECCS were also used for studying the deformation of PET-coatings during the DWI-process. These experiments are discussed in chapter 4.

For studying local differences of the corrosion activity in cans, a number of specialized techniques were also considered. These were the Scanning Reference Electrode Technique (SRET), the Scanning Vibrating Electrode Technique (SVET) and the Scanning Kelvin Probe (SKP). For all these techniques, the corroding surface is scanned from a close distance by a metal probe with a fine tip. The application of these techniques for corrosion research is relatively new and therefore in chapter 5, a general discussion on the interpretation of measurements by these techniques is given. In chapter 6, the SKP is studied in more detail.

Characterization of polymer coatings by dielectric spectroscopy can also be done for wider spectra than taken for the study in chapter 2. An interesting interpretation of such spectra is to consider them as the superposition of relaxation spectra with different time constants. This leads to the problem of the determination of time constant distributions from dielectric spectra, which problem can be handled by several different inversion algorithms. In chapter 7, three of such inversion algorithms are implemented and compared.



Chapter 2

Diffusion of water and ions through polymers and some measurements on PET

2.1 Introduction

In the introductory chapter 1 it was mentioned that for corrosion processes in polymer coated metal cans the transport of water and ions through the coating is important. As will be discussed, this transport is driven by concentration differences hence the transport is identical to diffusion. We will also discuss and compare several techniques for measuring the rate of water and ion diffusion in polymer coatings and we will present a number of actual diffusion measurements on free PET films and on PET coated ECCS.

2.2 Diffusion in polymers

As mentioned above, diffusion is the transport of particles driven by concentration differences. It turns out to be that many practical diffusion processes can be described by the same mathematical equations, which are generally referred to as Fick's laws. This holds for instance for:

- The diffusion of gases through a gas mixture;
- The diffusion of ions or molecules in solvents;
- The diffusion of neutrons through matter;
- The diffusion of ions and molecules through solids.

Of course the last case is most to our concern since it applies to the diffusion of water and ions through polymer coatings. For the discussion on diffusion mechanisms, it is however interesting to also keep the other examples in mind.

Below we will first discuss the main ideas behind Fick's laws and we will discuss under which conditions they are valid. Then we will discuss the diffusion of water and ions in polymer coatings within the framework of Fick's laws.

2.2.1 Fick's first law

Historically, Fick's laws were first written down by J.B.J. Fourier for the transport of heat through matter. Fourier also developed the basic solution methods for these equations [Fourier 1822]. The equations were taken over by A. Fick, who first used them for particle transport [Fick 1855]. At that time, the equations did not yet have a theoretical background and were based on the heuristic argument that the particle flow is opposite to the concentration gradient. Attempts to give Fick's laws a theoretical basis were for instance made by W. Nernst [Nernst 1888], but a rigorous treatment was first given by A. Einstein in one of his celebrated 1905 papers [Einstein 1905].

It is interesting to pay some attention to the physics behind Fick's laws, but the full derivation starting from Newton's laws on a microscopic scale is very lengthy. A much shorter and fully equivalent treatment can be given by starting with the chemical potential but then, the events on a molecular scale are largely ignored. As a compromise, we will only sketch the molecular treatment by a simple one-dimensional case and then move to the description with the chemical potential.

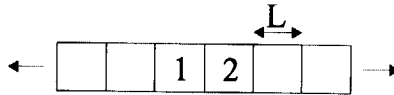


Figure 2.1. One dimensional diffusion in an infinitely long chain of cubes with length L . Two adjacent cubes are labeled by "1" and "2".

We first consider the diffusion through an infinite chain of adjacent cubes with length L , as illustrated by figure 2.1. Suppose that these cubes contain one species of particles and that the number of particles in block i is equal to N_i . Let us now suppose that in a time step τ , all particles make a random move to the adjacent cube. Since the move is random, half of the contents of each cube will go to the left cube while the other half goes to the right cube, on the average. This means that the net number of particles crossing the interface between two adjacent cubes 1 and 2 reads:

$$N_{1 \rightarrow 2} = \frac{1}{2} N_1 - \frac{1}{2} N_2$$

It follows that during the time step τ , there is a net transport from cube 1 to cube 2 when $N_1 > N_2$ and a net transport from 2 to 1 when $N_1 < N_2$. Hence, although the particle movement is essentially random, there is nevertheless a net transport from the higher to the lower concentrations. This is the origin of diffusion.

We will now write this simple formula into the usual form of Fick's first law and come to the definition of the diffusion coefficient for this particular case. First, we replace the number of particles N_i in cube i by the density $n_i = N_i / L^3$ in the cube:

$$N_{1 \rightarrow 2} = \frac{1}{2} n_1 L^3 - \frac{1}{2} n_2 L^3 = \frac{1}{2} L^3 (n_1 - n_2)$$

It is common to consider the net particle flow per unit of time and per unit of area where the particles are flowing through. This quantity is called the "particle current density" and is generally denoted by J . In our example, J reads:

$$J_{1 \rightarrow 2} \equiv \frac{N_{1 \rightarrow 2}}{\tau L^2} = \frac{1}{2} \frac{L}{\tau} (n_1 - n_2) \equiv \frac{1}{2} v (n_1 - n_2)$$

Here the ratio $v = L / \tau$ is defined, which is the magnitude of the random velocity of the particles. Now by considering n_i as a continuous function of the position x along the chain of cubes, one can write the following:

$$J(x) = -\frac{1}{2} v L \frac{dn}{dx} = -\frac{1}{2} \tau v^2 \frac{dn}{dx} \equiv -D \frac{dn}{dx}$$

This equation is Fick's first law and the quantity D is called the "diffusion coefficient". We see that the net particle flow is proportional to the opposite of the gradient of the particle concentration, in agreement with the original assumption of Fick in his 1855 paper.

This simple model can be refined to more realistic diffusion in 3 dimensions while keeping the main ideas. For instance, for the diffusion of ions in a solvent, the random velocity of the ions is connected to the thermal or kinetic energy of the ions, the mean time between collisions can be identified as τ and the mean free path as the mean distance L traveled in the time τ . The main differences between the analysis of the full three-dimensional case and our simple one-dimensional example are:

- Not only the random movement of the particles in the x -direction has to be considered, but also in the y and z -directions;
- Instead of fixed distances L in fixed time intervals τ , a distribution of τ and L values has to be considered.

When it is assumed that the probability of collision is independent of time, which is a very natural assumption, the distribution of the time between collisions is the Poisson distribution. With that and by also considering the particle movement in y and z directions, the following result can be obtained (cf. [Kittel 1976]):

$$D = \frac{1}{3} \tau v^2$$

Here, τ must be read as the mean time between collisions and v^2 as the mean square velocity. We see that this expression for D only differs by a constant factor from the expression for D

for our one-dimensional example. The mean square velocity is proportional to the temperature T by:

$$\frac{1}{2} m v^2 = \frac{3}{2} k_B T$$

where m is the mass per particle and $k_B = 1.4 \times 10^{-23}$ J/K is Boltzmann's constant. With that, the expression for D can also be written as:

$$D = \frac{\tau}{m} k_B T$$

This result holds for the diffusion of monatomic and mutually non-interacting particles. For diatomic particles or for molecules composed from even more atoms, the proportionality between v^2 and T still holds, but with another constant. This does not affect the main ideas of the discussion.

Diffusion and chemical potentials

Another path to Fick's first law is to consider the chemical potential μ of the particles, which reads (cf. [Kittel 1980]):

$$\mu = k_B T \ln(n/n_Q) + qV$$

Here, n_Q is the quantum concentration, q is the charge per particle and V is the electric potential. This expression holds for particles with no mutual interactions, such as ideal gases or ions in dilute solutions. The force F acting on a particle is minus the gradient of the chemical potential:

$$F = -\nabla\mu = -k_B T \frac{1}{n} \nabla n - q \nabla V$$

In the first term, the gradient of the particle density appears and therefore the first term is related to particle diffusion. The second term contains the gradient of the electric potential and governs the transport of electric charge. In thermal equilibrium, the particles will be distributed over the system such that the net force F on each particle in the system is zero. With that, one can for instance calculate things like the distribution of particles as a function of height in the gravitation field of the earth in thermal equilibrium.

For systems that are not in thermal equilibrium, for instance because an electric field is just applied or because some particles are just injected into the system, the net force F on the particles is nonzero and will lead to particle transport. In case the particles are moving through vacuum, F will lead to a constant acceleration by Newton's law. However when the particles are diffusing through a background material by which they are scattered within a mean time interval τ , the particles can only be accelerated over a distance L , which is their mean free path. Assuming that scattering randomizes the direction of the particles (this is the case when the background is rigid, or when the particles are much lighter than the background molecules, cf. [Duderstadt 1976]), the particles have to be accelerated again after

each collision in the direction of the force. By Newton's law, the acceleration is equal to F/m , where m is the mass per particle and it follows that in a time interval τ , a particle travels the following distance in the direction of the force:

$$L_{\text{drift}} = \frac{1}{2} \frac{F}{m} \tau^2$$

Considering a Poisson distribution for the time interval between collisions, one can derive the following for the mean drift velocity (cf. [Feynman 1964, Kittel 1976]):

$$v_{\text{drift}} = \frac{F\tau}{m}$$

where m is the mass per particle and τ is the mean time between collisions. It follows that for particle diffusion through a background medium, driven by a constant force F , the particles are not accelerated but transported with a constant mean drift velocity. The reason is that at every collision they lose energy to the background material, which is heated. One could also say that during their transport, they experience friction with the background material. (Note that the friction force is proportional to the drift velocity, which is characteristic for molecular friction).

In the absence of an electric field, the particles are driven by diffusion and one can write the following for the mean drift velocity:

$$v_{\text{drift}} = -\frac{\tau}{m} k_B T \frac{1}{n} \nabla n$$

Hence for diffusion through a background medium, the mean drift velocity of particle transport is proportional to the concentration gradient, which is the same result as found in the previous section. (It may seem strange to speak about a "force acting on the particles" in case of diffusion, since the particle movement is essentially random, but for gases the force can be interpreted as pressure.) The particle current density is just the product of the particle density and the mean drift velocity:

$$J = n v_{\text{drift}} = -\frac{\tau}{m} k_B T \nabla n = -D \nabla n$$

Thus we arrive at the same definition of D as in the previous section.

It is also interesting to consider the particle transport by an electric field. Assuming that the contribution of the relative concentration differences to the force F can be ignored, this leads to the following mean drift velocity:

$$v_{\text{drift}} = -\frac{\tau}{m} q \nabla V$$

and to the following for the electric current density:

$$J_{el} = qn v_{drift} = -\frac{\tau q^2 n}{m} \nabla V \equiv -\gamma \nabla V$$

This equation is nothing but Ohm's law, where γ is the electric conductivity. The ratio τ/m occurs both in D and in γ and is generally called the mobility of the particles in the given background medium. It follows that the mobility drops out when the ratio of D and in γ is taken:

$$\frac{D}{\gamma} = \frac{k_B T}{nq^2}$$

Hence under some well-defined conditions, D is a measure for γ and vice versa. This relation is known as the Stokes-Einstein relation.

2.2.2 Diffusion of water and ions in polymers

We have seen that for Fickian diffusion, the short-range collisions with a background material randomizing the particle movement and the absence of mutual interactions between the diffusing particles are characteristic. Now let us reconsider the four examples of diffusion mentioned earlier:

- Diffusion of gases through a gas mixture;
- Diffusion of ions or molecules in solvents;
- Diffusion of neutrons through matter;
- Diffusion of ions and molecules through solids.

In the first case, one can for instance think of Helium gas diffusing in the earth's atmosphere. The gas molecules of the atmosphere are then the background to which the Helium atoms are scattered. In the second case, the solvent molecules are the background and in the third case the atomic nuclei. We will now focus on the fourth case and especially to the diffusion of water and ions in polymers. Two leading mechanisms are generally suggested for the transport of molecules through polymers (cf. [Comyn 1985, Crank 1968]):

- Hopping of molecules between molecular sites. At these sites, the molecules are bonded to the polymer chains and these sites can be regarded as "potential pits";
- Transport through the "free volume" of the polymer. The free volume is a kind of inner cave structure of tunnels and rooms and originates from the production process of the polymer specimen.

Particle hopping is a completely different process than scattering, but can nevertheless equally lead to Fickian diffusion, since particle hopping also leads to random moves over a certain mean distance L in a mean time interval τ . In fact, the simple one-dimensional case discussed in section 2.1.1 is very close to particle hopping. Also for diffusion by particle hopping, energy is dissipated. This is the energy required to escape from a potential pit, at which action for instance a polymer chain has to be bent or displaced.

An interesting difference between diffusion by particle hopping and gas diffusion is however that for particle hopping, only a fraction $\exp(-A/k_B T)$ of the molecules has sufficient energy for hopping, where A is the activation energy. This means that the particle current due to hopping and hence the diffusion coefficient D are proportional to this factor $\exp(-A/k_B T)$, resulting in a completely different temperature dependence than for gas diffusion, as first mentioned by H. Braune [Braune 1924].

Particle hopping is the most likely process for the transport of particles that are strongly bonded to the background molecules. This will hold for ions that naturally will have a strong interaction with the background material due to their electric charge, but also for water molecules in case the background material is a hydrophilic polymer.

Also for the other transport mechanism for which molecules travel through the free volume of the polymer, Fick's laws may well be obeyed. When the free volume is a chaotic structure, it defines random paths for the diffusing molecules. Then the particles can also not be accelerated in one direction since their direction is randomized by curves in the tunnels, their speed is decreased in narrow passages and dead ends and so on.

Free volume transport is most likely for electrically neutral molecules such as gases. For hydrophobic polymers, the water will basically be transported through the free volume. It is an interesting question whether the water in the free volume can be considered liquid or not. It is suggested that if this is the case, the ion transport through the polymer could be facilitated, because the ions can then dissolve and be transported with the water (cf. [Vittel Rao 1978]).

For water diffusion in hydrophilic polymers, a combination of the two above mechanisms may occur. Then the diffusion of particles starts by filling the free volume followed by a deeper penetration by particle hopping. This leads to a typical water uptake behavior in two stages. Also a third phase is observed, in which the coating breaks down by the action of water [Wormwell 1949, Westing, van 1994]. The water diffusion is then characterized by two diffusion coefficients.

Apart from the fact that ions are normally much slower transported than water, their transport may also be slowed down by the fact that the positive ions have to wait for the negative ions and vice versa to maintain electric neutrality in space. For instance when the positive ions are more rapid, the diffusion front will become positively charged with respect to the tail. This results in a voltage along the diffusion path, which holds up the positive ions and drags the negative ions. This voltage is known as the "diffusion voltage" and is well known from the transport of ions through electrolytes (cf. [Vetter 1961]).

There are also cases of diffusion for which Fick's laws are not obeyed. One can think of the following reasons:

- The background medium is absent and hence there are no short range collisions. This holds for instance for diffusion in vacuum;
- The background material is composed from such light molecules (or atomic nuclei in case of neutron diffusion) that the direction of the diffusing particles is not completely randomized by a collision;
- The mutual interactions between the diffusing particles cannot be ignored. This holds for instance when the free volume is almost completely filled;

- Swelling of the polymer during water uptake. This is thought to be the explanation for so-called “Case II diffusion” [Alfrey 1966, Thomas 1982, Comyn 1985]. This type of diffusion is for strongly hydrophilic polymers.
- Different diffusion processes with different diffusion constants may play a role at one and the same time, such as particle hopping and free volume diffusion. In that case, Fick’s laws are actually not violated, but the diffusion will be different than expected from only a single D value. An elegant way to describe diffusion processes with more than one D value or even with a continuous spectrum of D values is given by Wind and Lenderink [Wind 1996];
- Another reason why a single D value may not be sufficient may be the spatial variation of D , as in polymer coatings that are build up from different layers. Alternatively, the structure of the free volume may not be isotropic. Then D will no be a scalar but a tensor, giving the relations between the components of J and the gradient of n .

This is by no means a complete list and the reasons and mechanisms for “non-Fickian diffusion” are still subject of discussion.

2.2.3 The diffusion equation, Fick’s second law

Fick’s first law, stating that the particle current density is proportional to the gradient of the particle concentration, is the most significant part of Fickian diffusion. However, its form with the particle current density:

$$J = -D\nabla n$$

is not practical in case one wants to describe the concentration of particles as a function of the position r and time t during a diffusion process. In that case, one is only interested in $n(r,t)$ and not in $J(r,t)$. The common way to eliminate the particle current density J is to apply Gauss’s law:

$$\nabla \cdot J = -\frac{\partial n}{\partial t}$$

This equation is nothing but the conservation of particles. It says that for any volume element, the net flow of particles out of the volume element (which is the divergence of J) is equal to the decrease of particles in the volume element with time. Substitution of Fick’s first law leads to the following:

$$\frac{\partial n}{\partial t} = \nabla \cdot (D\nabla n)$$

This partial differential equation for the particle density $n(r,t)$ is known as the diffusion equation or also as Fick’s second law.

The diffusion equation is, together with the wave equation, the best known partial differential equation. In the following section, we will discuss the solution of the diffusion equation for the diffusion of water and ions in polymer coatings.

2.2.4 Diffusion of water and ions in polymer films

The main solution method of the diffusion equation can already be found in the pioneering work of Fourier. The solution method involves expansion of the solution $n(r,t)$ into a series of eigenfunctions and is nowadays a well-established technique, which is well explained in many textbooks (cf. [Crank 1968, Boyce 1977]). Here we will pay no attention to the solution method and just briefly discuss the solutions for the diffusion in polymer films.

Assuming that the particle concentration in the film only varies in the direction perpendicularly to the surface of the film, say the x -direction, Fick's second law for the particle density simplifies to:

$$\frac{\partial n}{\partial t} = D \frac{\partial^2 n}{\partial x^2}$$

We have also assumed that the polymer film is homogeneous, hence that D is independent of position. As we will see, all solutions relevant to this work can be derived from the diffusion of particles out of a film that is initially, at $t = 0$, filled with a homogeneous particle concentration n_0 . For a film with thickness L , the solution for $n(x,t)$ reads:

$$n(x,t) = \sum_{\text{odd}(k)} \frac{4n_0}{k\pi} \exp\left[-\frac{k^2\pi^2 D}{L^2} t\right] \sin\left(\frac{k\pi}{L} x\right) \quad (2.1)$$

In this expression, the dummy variable k runs over all odd numbers 1, 3, 5, and so on. The solution is valid between $x = 0$ and $x = L$, where the film is located.

At $t = 0$, all exponentials are equal to 1 and the solution is the Fourier series of a square wave with period $2L$, with $n(x,t) = n_0$ for x between 0 and L and with $n(x,t) = 0$ for x between L and $2L$. Since the solution is only valid between $x = 0$ and L , the only relevant part implies that n is equal to the initial concentration n_0 of the homogeneously filled film, in agreement with the initial condition. For longer times t , the exponentials more and more decrease and eventually n will be zero after an infinitely long time, corresponding to the final state where all particles have left the film. A more precise estimate of the long term diffusion out of the film is obtained from the consideration that for larger k values, the exponentials decay more rapidly. Hence for longer times, the term with $k = 1$ will eventually become dominant, which leads to the following approximation for the long term behavior:

$$n(x,t \rightarrow \infty) \cong \frac{4n_0}{\pi} \exp\left[-\frac{\pi^2 D}{L^2} t\right] \sin\left(\frac{\pi}{L} x\right)$$

We see that for longer times, the shape of n is a sine with exponentially decreasing amplitude. In figure 2.2, the solution is plotted for different times.

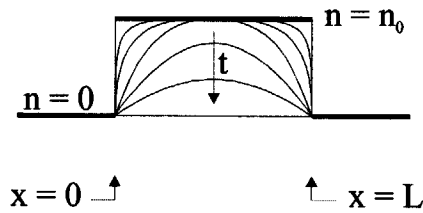


Figure 2.2. Diffusion of particles out of a free film. Initially, the film is homogeneously filled with a concentration n_0 and eventually all particles have left the film.

For the evaporation of a gas out of a polymer film, it is assumed that as soon as the particles have left the film, they are rapidly carried away, such as by a dry air stream in case of water. Then the particle concentration on the boundaries of the film and outside the film can be considered zero and the medium outside the film has no influence on the diffusion process.

This is different for the opposite diffusion problem where particles are diffusing into a film from a medium. In that case, one has also to deal with the problem how the particle concentrations inside and outside the films are related. The easiest solution to this problem is to assume that around the interface between the polymer and the medium, the particles are in thermal equilibrium. Then the chemical potentials in the polymer and the medium are equal:

$$\mu^{\text{pol}} = k_B T \ln \frac{n^{\text{pol}}}{n_Q^{\text{pol}}} = \mu^{\text{med}} = k_B T \ln \frac{n^{\text{med}}}{n_Q^{\text{med}}}$$

where n^{pol} and n^{med} are the particle concentrations in the polymer and the medium and n_Q^{pol} and n_Q^{med} are the quantum concentrations. It follows that:

$$n^{\text{pol}} = \frac{n_Q^{\text{pol}}}{n_Q^{\text{med}}} n^{\text{med}} \equiv \beta n^{\text{med}}$$

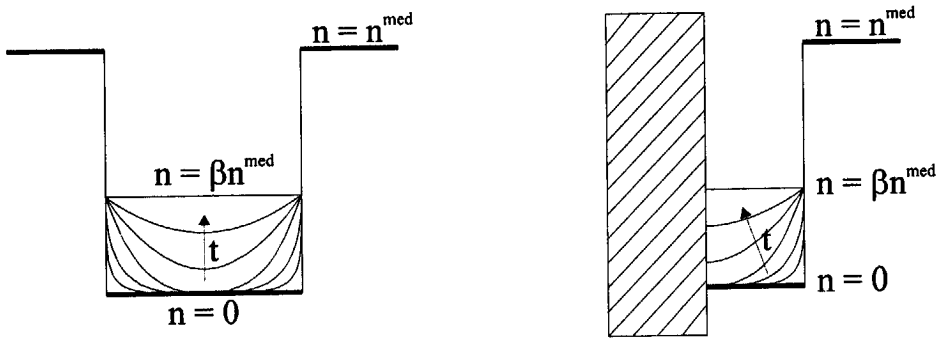
Hence around the interface, the particle concentrations in the polymer and the medium only differ by a constant factor. This is Henry's law (cf. [Comyn 1985]). Note that it is assumed that the mutual interactions of the diffusing particles in the polymer can be ignored, otherwise the above expression for μ^{pol} is not valid.

The proportionality constant β is called the solubility of the particles from the medium into the polymer. (Usually the symbol S is used for the solubility, but in this chapter, we will use S for the conductance of the coating. In fact, the solubility is defined in different ways, not seldomly leading to non-SI units. With to the above definition, β is dimensionless.)

For the uptake of particles from a medium, it is generally assumed that at the interface, the particles are rapidly supplied from the medium, such that the bulk concentration in the medium holds until the polymer surface. In addition, it is assumed that the interface is in

thermal equilibrium, hence that Henry's law is obeyed for the particle concentrations around the interface. Then, the particle concentration at the polymer side of the interface is right from the start equal to the saturation concentration βn^{med} . Under these conditions, the diffusion from the medium into the film is simply the saturation concentration βn^{med} minus equation (2.1) with n_0 equal to βn^{med} . This solution is plotted in figure 2.3.

For a supported film, such as a polymer coating on a metal, the particles only enter the film from one side and the boundary conditions of the diffusion problem are different. However from symmetry considerations, it follows that the solution can be found from taking only half of the solution for a free film that is two times thicker, as shown in figures 2.3 and 2.4.



Figures 2.3 and 2.4. The uptake of particles for a free and for a supported film. The solution for the supported film is just the right part of the solution for the free film.

2.3 Diffusion measurements

The transport of particles through a polymer film is characterized by the diffusion coefficient D and the solubility β of the particles in the polymer. In fact one can show that the product $D\beta$, which is called the "permeability", is a good measure for the amount of particles transported through the film under given conditions. This is illustrated by the following experiment where a polymer film with thickness L is mounted between two gas chambers with different particle densities n_1 and n_2 as shown in figure 2.5. The particle densities are maintained by keeping the (partial) pressures of the diffusing gas in the chambers constant. In the steady state, the following particle current density will flow through the film:

$$J = D \frac{n_1^{\text{pol}} - n_2^{\text{pol}}}{L} = D\beta \frac{n_1 - n_2}{L}$$

where it is assumed that Henry's law is obeyed. It follows that for a fixed gas concentration difference between the chambers, the product $D\beta$ is a measure for J and hence for the particle transport through the film.

By such experiments, which are known as permeability measurements, the product $D\beta$ is determined from a measurement of J for given particle densities n_1 and n_2 . During the experiment, n_1 and n_2 are fixed by maintaining a constant (partial) pressure in the chambers. Not only the permeability $D\beta$, but also the separate value of D (and with that the separate value of β) can be determined by measuring how much time it takes to reach a steady state flow (cf. [Crank 1968, Comyn 1985]).

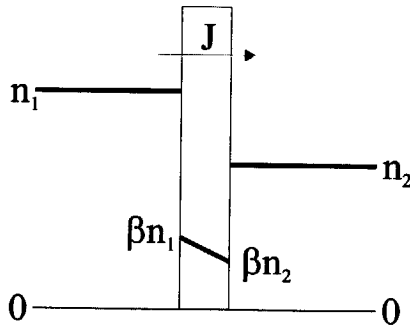


Figure 2.5. Permeability measurements. The particle density on the left-hand side is higher than on the right-hand side. The stationary particle densities are plotted.

Permeability measurements are normally done for gas diffusion. The diffusion of water is usually measured with water vapor. One may wonder whether the diffusion from water vapor and from liquid water are comparable, since the concentration of water molecules in liquid water is much higher than in water vapor. Especially when one pictures the water transport through a polymer as the penetration of the inner cave structure of the free volume, one would expect that the uptake from liquid water is much higher than from water vapor.

In practice however, it is generally observed that the total uptake from 100% saturated water vapor and from liquid water are equal (cf. [Comyn 1985]). This even holds for hydrophobic polymers such as PET where the water is basically transported through the free volume [Yasuda 1962]. This suggests that the distinction between liquid water and water vapor does not exist for the water inside a polymer. From this point of view it makes sense that the uptake from 100% saturated water vapor and from liquid water are equal, since 100% saturated water vapor and liquid water are in thermodynamic equilibrium (or in other words, their chemical potentials are equal). This makes the idea that the ion transport could be accelerated by the dissolution of ions in the water of the free volume unlikely, since in general the free volume is apparently too narrow for the formation of liquid water.

Permeability measurements can also be done for ion diffusion measurements by filling the chambers with electrolytes with different ion concentrations. Then one has to follow the ion concentrations in the chambers as a function of time, which can for instance be done for Cl^- ions by using a calomel electrode or a silver chloride electrode [Kittelberger 1952, Murray 1973]. Also the transport of H^+ could be followed by using a pH electrode. Alternatively, one can also consider working with radioisotopes such as ^{22}Na and ^{36}Cl in order to follow the ion transport [Glass 1967].

A drawback of permeability measurements is that they only can be done with free films. In our case, we are basically interested in the diffusion of water and ions in PET films on ECCS, hence in supported coatings. Free PET films of the same composition were also available, but since the production method of free and supported films is essentially different, it cannot be excluded that the diffusion properties of free and supported films are different. Therefore we considered two different techniques of diffusion measurements by which the diffusion in supported films can be studied:

- Gravimetric measurements, by which the mass changes of the polymer coating are followed during diffusion;
- Impedance measurements, for which the permittivity and the conductance of the coating are followed.

These techniques are discussed in the following sections. Gravimetric measurements are well known as a technique for diffusion measurements and they are also described in standard works on diffusion in polymers (cf. [Crank 1968, Comyn 1985]). Impedance measurements are only applicable for the uptake of water and ions from electrolytes, which is probably the reason that they are not widely known outside the research area of corrosion protective coatings. The technique is however not very new [Wormwell 1949].

We will see that gravimetric measurements are especially suitable for determining the total amount of particles taken up in the film (hence for determining the solubility β), while impedance measurements are more sensitive for determining the mobility of water and ions in the film (hence for determining the diffusion coefficient D). The techniques are therefore complementary.

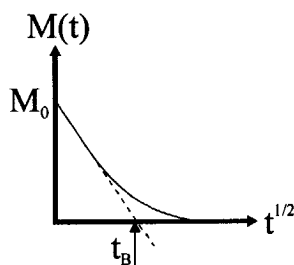


Figure 2.6. Total mass of particles in a free polymer film when the particles leave the film by diffusion. The total mass is plotted as a function of the square root of time.

2.3.1 Gravimetric measurements

For this technique, the amount of particles diffusing in or out of a polymer film is followed by measuring the total mass of the film by a microbalance. We will first consider the diffusion of particles out of a free polymer film. Assuming that Fick's laws are obeyed, the total mass of the particles in the film as a function of time can be calculated by integration of equation (2.1) over the whole thickness of the film. This leads to the following:

$$M(t) = M_0 \sum_{\text{odd}(k)} \frac{8}{k^2 \pi^2} \exp\left[-\frac{k^2 \pi^2 D}{L^2} t\right] \quad (2.2)$$

where M_0 is the initial mass of the particles in the film. The function $M(t)$ is plotted in figure 2.6. With equation (2.2), the diffusion coefficient D can be determined from a measurement of $M(t)$ by a fitting procedure. The exact solution for $M(t)$ from Fick's laws is however rather complicated and therefore one usually considers the limiting behavior of the exact solution. One possibility is to consider the long-term behavior, for which the term with $k = 1$ becomes dominant and the total mass of the particles in the film decays exponentially. However, the mass differences for larger times are relatively small and therefore one rather determines D from the short-term behavior of $M(t)$. An approximation for $M(t)$ for short times follows by first differentiating $M(t)$ to the time:

$$\frac{dM}{dt} = -\frac{8M_0 D}{L^2} \sum_{\text{odd}(k)} \exp\left[-\frac{k^2 \pi^2 D}{L^2} t\right]$$

By this differentiation, one eliminates the factor $1/k^2$ in the summation. Now for small t , the terms with consecutive k are not much different and the summation over k can be approximated by a well-known elementary integral, giving:

$$\frac{dM}{dt} \cong -\frac{8M_0 D}{L^2} \frac{1}{2} \int_{k=0}^{\infty} \exp\left[-\frac{k^2 \pi^2 D}{L^2} t\right] dk = -\frac{2M_0}{L\sqrt{\pi Dt}}$$

Note the extra factor $1/2$ that accounts for the fact that the summation is only carried out over the odd k values. Integration to the time and setting the integration constant to M_0 finally gives the following approximation for short times:

$$M(t \rightarrow 0) \cong M_0 \left(1 - \frac{4}{L} \sqrt{\frac{Dt}{\pi}}\right)$$

Hence when M is plotted versus the square root of t , the result is a straight line for short times, and D follows from the slope of this line. Equivalently, the asymptote for short times can be extrapolated to the $M = 0$ axis. The intersection defines a "breakpoint time" t_B , which is a measure for the duration of the particle diffusion out of the film. This procedure is illustrated by figure 2.6 and leads to the following useful formula for the diffusion coefficient as a function of the experimentally obtainable breakpoint time t_B :

$$D = \frac{\pi L^2}{16 t_B}$$

For the inverse process of the uptake of an empty film, a similar method can be followed. Then, the asymptotic behavior of $M(t)$ for short times reads:

$$M(t \rightarrow 0) \cong M_{\text{sat}} \frac{4}{L} \sqrt{\frac{Dt}{\pi}}$$

where M_{sat} is the total mass of the particles when the film is saturated. In this case, the breakpoint time can be defined as the intersection with the horizontal line $M = M_{\text{sat}}$, which leads to the same expression for the breakpoint time and hence to the same formula for the diffusion coefficient.

For the diffusion in or out of supported films of thickness L , the same methods can be used, but then one should replace L by $2L$ in the above expressions, as discussed above.

Gravimetric measurements are well suited for determining the solubility β , since the excess mass of the polymer film is simply proportional to the total amount of particles taken up by the film.

In principle, gravimetric measurements can both be used for free and supported polymer films. However for supported films, one must be very sure that during the diffusion process, the mass of the supporting material underneath the film does not change. The latter may not only happen due to the diffusion of particles into the supporting material, but also due to corrosion in case the supporting material is steel.

2.3.2 Electric impedance measurements

Another idea of measuring the diffusion of water and ions in polymer coatings is to measure the electric impedance of the coating. In general, the electric impedance of polymers is strongly affected by a water or ion content because of the large permanent electric dipole moment of water molecules and because the electric conductivity of most polymers is largely determined by the ion content.

The standard experimental setup for diffusion measurements with the electric impedance is sketched in figure 2.7. In this setup, a plane sample of polymer coated material is immersed by an electrolyte and the electric impedance between the metal and the electrolyte is measured as a function of time. Thus, the electric impedance of a supported coating is followed during the water and ion uptake from an electrolyte.

In chapter 3, more will be said about the concept of the electric impedance and the associated measurement technique. Here, we take the expression for the electric impedance Z of a polymer coating as a starting point. For our purposes, it is most convenient to consider the inverse of Z that reads:

$$\frac{1}{Z} = S + i\omega C$$

where S is the conductance of the coating (which is the inverse of the resistance), C is the capacitance, i is the imaginary unit and ω is the angular frequency at which $1/Z$ is measured. In fact, the coating can be regarded as a network of a parallel resistor and capacitor. The quantities S and C can be expressed by the material properties γ and ϵ , which are the electric conductivity and the permittivity of the polymer (cf. [Böttcher 1978]). This yields:

$$\frac{1}{Z} = \frac{\gamma A}{L} + i\omega \frac{\epsilon A}{L} = \frac{A}{L} (\gamma + i\omega\epsilon)$$

where L and A are the thickness and the area of the coating. Hence, the electric impedance Z of the coating contains contributions from the electric conductivity and the permittivity of the coating material.

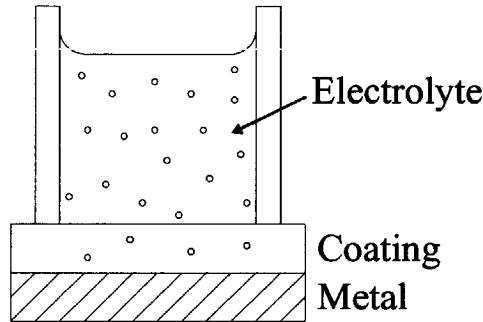


Figure 2.7. *Experimental setup for the uptake of water and ions for a polymer coating on metal. The electric impedance between the electrolyte and the metal is measured.*

Now the key idea is that γ is a measure for the ion uptake and ϵ is a measure of the water uptake. Hence one could think of measuring the water and ion uptake separately by measuring the real and imaginary parts of $1/Z$. However, the following considerations have to be taken into account:

- During uptake, the water and ions are not homogeneously distributed in the coating, but obey a diffusion profile such as equation (2.1);
- The conductivity and the permittivity are complex quantities and may contain imaginary parts, depending on the angular frequency ω . As a result, the real part of $1/Z$ may contain a contribution of ϵ and the imaginary part of $1/Z$ may contain a contribution of γ ;
- What is the relation between the observed differences in γ and ϵ and the actual quantities of ions and water that have been taken up?

In case the water and ions are distributed in layers parallel to the coating surface, the total impedance of the coating Z is obtained by a summation over the impedances of the layers of the coating:

$$Z = \int_{x=0}^L dZ(x) = \frac{1}{A} \int_{x=0}^L \frac{dx}{\gamma(x) + i\omega\epsilon(x)}$$

It follows that γ and ϵ are entangled in the integral and cannot be separated unless some kind of approximation is made. For that, we observe that the latter expression for Z contains the harmonic mean of $\gamma + i\omega\epsilon$:

$$Z = \frac{L}{A} \frac{1}{\langle \gamma + i\omega\epsilon \rangle_H}$$

Now in case the real and imaginary parts of $\gamma + i\omega\epsilon$ do not vary much over the coating, the harmonic mean is approximately equal to the arithmetic mean and one can write:

$$Z = \frac{L}{A} \frac{1}{\langle \gamma + i\omega\epsilon \rangle_H} \cong \frac{L}{A} \frac{1}{\langle \gamma + i\omega\epsilon \rangle_A}$$

which leads to the following approximation for $1/Z$:

$$\frac{1}{Z} \cong \frac{A}{L} \langle \gamma + i\omega\epsilon \rangle_A = \frac{A}{L} \frac{1}{L} \int_{x=0}^L [\gamma(x) + i\omega\epsilon(x)] dx$$

By this approximation, the contributions of the real and imaginary parts of $\gamma + i\omega\epsilon$ can be separated. This does not automatically mean that γ and ϵ can be separated, since both γ and ϵ may contain an imaginary part, as remarked above.

For water uptake measurements, the frequency of the impedance measurement is normally below 1 MHz because in this frequency range, the contribution of permanent dipoles to the permittivity is maximal and the effect of water is most significant. At such low frequencies, γ can be considered real and independent of frequency but ϵ will generally have a small but considerable imaginary part. It is common to separate the real and imaginary parts of ϵ by:

$$\epsilon = \epsilon' - i\epsilon''$$

(cf. [Böttcher 1978]). Substitution in the above expression for $1/Z$ gives:

$$\frac{1}{Z} \cong \frac{A}{L} \frac{1}{L} \int_{x=0}^L [\gamma(x) + \omega\epsilon''(x) + i\omega\epsilon'(x)] dx$$

This leads to the following approximations for the conductance and the capacitance of the coating:

$$S \cong \frac{A}{L} \frac{1}{L} \int_{x=0}^L [\gamma(x) + \omega\epsilon''(x)] dx = \frac{A}{L} \langle \gamma + \omega\epsilon'' \rangle_A$$

$$C \cong \frac{A}{L} \frac{1}{L} \int_{x=0}^L \epsilon'(x) dx = \frac{A}{L} \langle \epsilon' \rangle_A$$

The quantities S and C can be obtained from a measurement of $1/Z$. We see that the C only contains ϵ' and can therefore be used for water uptake measurements. The conductivity S contains both γ and ϵ'' and is therefore affected by both the ion and water uptake.

The approximations for S and C are only valid when the harmonic means of $\gamma + \omega\epsilon''$ and ϵ' can be approximated by the arithmetic means, which is the case when these quantities do not vary much over the coating. This will mostly hold for ϵ' and ϵ'' since the water content mostly changes the permittivity of the coating by no more than a few percent. Hence, the approximation for C will normally hold and since for polymers, S is usually dominated by the contribution $\omega\epsilon''$, the approximation for S will mostly also be valid. (For very low frequencies however, S will eventually be dominated by γ . Then the approximation for S may be very bad since γ of most polymers is extremely dependent on the ion content. For instance, when the ion front has penetrated half of the coating, γ is zero at the empty layers and the overall S of the coating is zero too. This is correctly described by the harmonic mean, but not by the arithmetic mean, which is therefore no good approximation in that case.)

By following the change of S and C as a function of time, the diffusion constants of water and ions can be determined by the breakpoint time method, similar to the gravimetric uptake measurements. For also determining the solubility, it is necessary to translate the changes of S and C to actual ion and water contents. This is by no means trivial and in fact, no satisfactory solution to this problem has been found yet. In the remaining part of this subsection this issue will be discussed.

Water uptake measurements

The idea of using the capacitance of a polymer coating as a measure for the amount of water taken up by the polymer coating dates from the work of Hartshorn, Megson and Rushton [Hartshorn 1937]. Several years after this work, Wormwell and Brasher published a study in which C was followed as a function of time for various polymer coatings, from which measurements they could determine the diffusion coefficient of water transport in the coatings [Wormwell 1949]. In this pioneering study it was also shown that the capacitance of the coating typically increases in three stages:

- A rapid increase due to the water uptake in the free volume of the polymer;
- A very slow increase by water uptake in the polymer structure;
- An eventual increase when the polymer structure breaks down.

Again a few years later, Brasher and Kingsbury compared the increase of the capacitance with gravimetric measurements [Brasher 1954]. After the latter paper, capacitive water uptake studies have become increasingly popular for studying corrosion protective coatings.

As mentioned above, a major difficulty of the interpretation of capacitive water uptake measurements is how the permittivity of a polymer changes as when water enters the polymer structure. A simple pragmatic approach is to assume that:

- All water molecules have the same impact on ϵ' of the polymer;
- The change in ϵ' of the polymer by water uptake is relatively small.

In that case, C will increase linearly with the amount of water taken up and that one can write:

$$\varepsilon'(v_w) = \varepsilon'_p + \lambda v_w$$

where v_w is the volume fraction of water in the polymer structure, ε'_p is the real part of the permittivity for the polymer without water (hence with $v_w = 0$) and λ is a proportionality constant ($\lambda = d\varepsilon'/dv_w$ at $v_w = 0$).

With this linear mixing formula, one can calculate how the capacitance of a polymer film increases with time during water diffusion. Assuming that Fick's laws are obeyed for a single D value, one can for instance obtain the following for the water uptake of a free polymer film:

$$\frac{C(t) - C_0}{C_0} = 1 - \sum_{\text{odd}(k)} \frac{8}{k^2 \pi^2} \exp\left[-\frac{k^2 \pi^2 D}{L^2} t\right] \quad (2.3)$$

where C_0 is the initial capacitance at $t = 0$. The right hand side is exactly equal to the relative mass increase of the film. This is not surprising, because the mass increase of the film is also linear with the amount of water taken up. For a supported film, the expression has as always to be modified by replacing L by $2L$.

When a measurement of $C(t)$ resembles equation (2.3), the water diffusion coefficient can be determined by a fitting procedure similar to gravimetric diffusion measurements. The total amount of water however does not follow, since the value of λ will not be known in general. Hence, in contrast with gravimetric diffusion measurements, the solubility and the permeability of water do not follow.

This is a serious drawback from capacitive diffusion measurements and efforts have been spent to determine the permittivity of a mixture of dielectrics from theoretical considerations. This however turns out to be very difficult for several reasons. In fact already for a homogeneous dielectric, the relation between the electric dipole moment of the molecules and the macroscopic permittivity is a problem too difficult for analytical methods and the case of a mixture of two dielectrics, such as a polymer with water, is even more complicated. The difficulty is that the local electric field acting on the dipoles is not simply equal to the externally applied field, but is also influenced by the dipoles in the neighborhood by some unknown fraction. Only for very diluted dielectrics such as gases, a satisfactory analytical treatment can be given because then, the mutual interaction between the dipoles can be ignored.

Nevertheless, when the permittivities of the different components of a mixture of dielectrics are known, one can make a guess of the permittivity of the mixture by taking some kind of average of the permittivities of the components. For that purpose, a number of mixture formulas have been proposed and some of the most popular ones can be written in the following form with a parameter α (cf. [Wind 1996]):

$$\varepsilon^\alpha = v_1 \varepsilon_1^\alpha + v_2 \varepsilon_2^\alpha$$

Here ϵ is the average dielectric constant of the mixture and ϵ_1 and ϵ_2 are the dielectric constants of the components. Further, v_1 and v_2 are the volume-fractions of the components in the mixture, which add up to 1. The mixture formula is used for both the complex permittivity ϵ as well as for the real part ϵ' only. (In general this does not lead to the same result but in our case, where ϵ' is much larger than ϵ'' , the difference is very small.)

The parameter α depends on the way the two components are mixed. Two values of α , the values -1 and $+1$, are supported by a solid physical interpretation. The value $\alpha = -1$ corresponds to a pile of water and polymer layers, parallel to the coating surface. In that case, the electric field is perpendicular to the polymer-water interfaces, leading to the harmonic mean of ϵ_1 and ϵ_2 . The value $\alpha = +1$ on the other hand corresponds to the case where the polymer-water interfaces are parallel with the applied electric field. The latter holds for instance for a coating with vertical pores filled with water.

In case the water is randomly distributed in a sort of cave structure, neither of the cases $\alpha = -1$ or $+1$ will apply. In this case, Hartshorn, Megson and Rushton proposed to use the average value $\alpha = 0$ [Hartshorn 1937]. Since with $\alpha = 0$, the mixture formula comes down to the triviality $1 = v_1 + v_2$, the limiting case of α near zero has to be considered. By writing $\epsilon^\alpha = \exp(\alpha \ln \epsilon)$, which approaches $1 + \alpha \ln \epsilon$ for α near zero, the following is obtained:

$$\ln(\epsilon) = v_1 \ln(\epsilon_1) + v_2 \ln(\epsilon_2)$$

This logarithmic mixture formula, first written by Hartshorn, Megson and Rushton [Hartshorn 1937], it is commonly known as the Brasher-Kingsbury mixture formula after Brasher and Kingsbury who also used it [Brasher 1954].

The logarithmic mixture formula is the most commonly used formula in the research of corrosion protective coatings and in spite of its "ad hoc derivation" it seems in better agreement with gravimetric uptake measurements than other mixture formulas for many types of coatings [Lindqvist 1985]. However in similar comparative studies, inconsistencies are also regularly reported (cf. [Brasher 1954, Westing van 1994]). These could be explained by the following reasons:

- The inhomogeneous water distribution in the coating, by which the expression for C as an integral over ϵ' only holds as an approximation. This affects the shape of the curve $C(t)$ during water uptake;
- The coating itself can be inhomogeneous. This can be unintentionally the case, for instance when the coating has dried up inhomogeneously after application, but also because the coating is build up from different polymer layers;
- The water molecules in the coating may be hindered in their movement in case they are bonded to the polymer molecules. This will be the case for hydrophilic polymers. Depending on how much the water molecules are hindered in their movement, the relative permittivity of the water appearing in the mixture formula, may range between 80 (for free water) and 1 (for absolutely immobile water) and there is no way of telling which value one should take;
- The water molecules may appear both in the free volume of the coating and bond to the polymer molecules. Then not only two different values of the diffusion coefficient have to

be used to describe the water diffusion, but also two different values of the permittivity of the water taken up;

- Vertical pores are perhaps filled more rapidly with water than horizontal pores. Then in the beginning, the parameter α of the mixture formula will be closer to +1, which means that the first water molecules entering the coating have a relatively large impact on the capacitance. This argument was given by Brasher and Kingsbury in order to explain their observation that for several coatings, the capacitance $C(t)$ initially increased more rapidly than the mass $M(t)$ [Brasher 1954];
- Swelling of the coating during water uptake. This particularly holds for hydrophilic coatings. One can think of correcting C for the swelling by measuring the thickness of the coating L [Brasher 1954];
- Finally, the validity of gravimetric measurements is also not absolute. For instance, an undesired increase of the mass can occur due to water sticking on the surface. For supported coatings, the supporting material may also take up water and in case the supporting material is a metal, it may change mass due to corrosion.

It follows that determining the actual amount of water in a coating from capacitive diffusion measurements is cumbersome.

Ion uptake measurements

The conductance S of a polymer coating both contains contributions from the conductivity γ and from the imaginary part ϵ'' of the permittivity. Both γ and ϵ'' are due to do friction losses: γ is a measure for the friction losses during ion transport and ϵ'' is a measure for the friction losses when electric dipoles are rotated. For the latter reason, ϵ'' is often commonly known as the "dielectric loss".

Under immersion of an electrolyte, S is affected by both the ion and water uptake, since γ is affected by the ion uptake and ϵ'' by the water uptake. Since the water uptake can already be studied by the capacitance C of the coating, S is most interesting for ion uptake measurements. But then, the contribution of ϵ'' must be eliminated. One can think of the following ways to achieve this:

- Measure the electric impedance Z of the coating at zero frequency;
- Measure an impedance spectrum $Z(\omega)$ for a set of ω values. Then in case $\epsilon'(\omega)$ shows little variation with frequency, one may assume that $\epsilon''(\omega)$ also shows little variation with frequency and $Z(0)$ can be obtained by fitting a straight line through the real part of $1/Z(\omega)$. In case $\epsilon'(\omega)$ significantly varies with frequency, one may calculate $\epsilon''(\omega)$ from $\epsilon'(\omega)$ by the Kramers-Kronig transform (cf. [Böttcher 1978, Willems 1995]);
- In case the ion transport is much slower than the water transport, S will first show a rapid increase by the water uptake and then a slower increase due to the ion uptake. One can then extract the contribution of γ by ignoring the initial rapid increase of S , possibly by comparison with the capacitance C , which also contains the water uptake.

From theoretical considerations, the first option may seem the most obvious. However, small currents of lower frequencies are very difficult to measure accurately, since it takes a very long time to collect data over a large number of periods. In addition, applying a direct voltage

over the coating will significantly enhance the ion transport into the coating and disturbs the diffusive ion uptake that one wants to measure. Only when the DC conductance of the coating is relatively high and the ion transport is relatively fast, DC measurements of Z may be an option.

Measuring an impedance spectrum for a considerable number of frequencies is more time consuming than only measuring Z for a single frequency, but may be the best way to obtain γ from S in the general case when the ion transport is not necessarily much slower than the water transport.

When γ is obtained as a function of time, the diffusion coefficient of the ions can be determined from the breakpoint time, similar to the water uptake measurements. For the determination of the solubility of the ions, γ must be related to the ion density. This is more obvious than finding a relation between ϵ' and the water density, since the conductivity γ of a polymer is simply proportional to the ion concentration, as discussed in section 2.1.2:

$$\gamma = \frac{\tau q^2 n}{m}$$

This does however not mean that the ion concentration n follows immediately from γ since in general, we will not know the value of the mobility τ/m of the ions in the polymer. In addition, when the ion transport is governed by ion hopping, n is only the fraction of the ions with sufficient energy for crossing the potential barrier. Nevertheless it is fair to state that at a fixed temperature, γ and n are proportional and that γ is a good measure for the ion uptake.

Another idea is to derive the ion diffusion constant D from the ion conductivity γ via the Stokes-Einstein relation (cf. [Bacon 1948, Kittelberger 1952]):

$$D = \frac{k_B T}{nq^2} \gamma$$

This equation relies on the fact that both D and γ are proportional to the mobility (see section 2.2.1). A practical problem is however that in our case, the density n of mobile particles is unknown. In addition, γ may be no good measure for D in case the mobilities of positive and negative ions are different. Then the electric current will be dominated by the more mobile ions (positive or negative), while for the diffusion current the positive and negative ions must travel together for maintaining electric neutrality. (Here the diffusion voltage, mentioned in section 2.2.2, might be recalled.)

2.4 Water and ion diffusion measurements on PET

Around 1950, PET was already produced commercially under the trade name Mylar by DuPont and around 1960, several studies on the diffusion properties of PET had already been published. More recently, the application of PET for bottles with carbonated beverages has lead to new research on the transport properties of PET, especially for the diffusion of oxygen

and carbon dioxide. In section 2.4.1, a short survey on previously obtained results on the transport properties of PET is given, focusing on the transport of water and ions.

The transport properties of polymers strongly depend on the production process and therefore we did a number of diffusion measurements on our own PET films. In addition we studied the influence of certain environmental parameters on the diffusion of water and ions in PET. We both did gravimetric and electric impedance measurements and the results are discussed in sections 2.4.2, 2.4.3 and 2.4.4.

2.4.1 Results on water and ion diffusion in PET from literature

Because of its hydrophobic behavior, it is generally thought that water molecules inside PET are not bonded to the polymer chains and that water in PET is only transported through the free volume (cf. [Lasoski 1959]). The same holds for many gases that have no interaction with the PET chains (cf. [Myers 1957, Maeres 1958, Michaels 1963]).

Depending on the production method, PET partly consists of amorphous and crystalline regions. The percentage of crystallinity usually ranges from zero for entirely amorphous PET to about 50%. The density of crystalline PET is 1.455 g/cm^3 , which is considerably higher than the density of amorphous PET, which is 1.330 g/cm^3 [Lasoski 1959]. It is thought that the free volume of a PET specimen is basically present in the amorphous region and that the percentage of crystallinity is therefore an important parameter for the transport properties of PET.

Amorphous PET can be obtained by quenching PET from the melt. This can for instance be done at a rate of 40°C per minute [Willems 1995]. Thereafter, the degree of crystallinity can be increased up to about 50% by an annealing procedure, for which the PET is heated to temperatures between 110 and 130°C [Michaels 1963, McCrum 1967, Willems 1995]. For instance, after 1 hour of annealing at 120°C , a crystallinity of 35% is obtained [Michaels 1963]. The degree of crystallinity of a PET specimen can be determined by measuring the density (cf. [Lasoski 1959]) or by differential thermal analysis (cf. [Michaels 1963, Willems 1995]).

For water diffusion in amorphous PET films, Yasuda and Stannett found by permeability measurements on free films the values $D = 3.94 \times 10^{-9} \text{ cm}^2/\text{s}$ for the diffusion coefficient and $\beta = 11.1$ for the solubility at room temperature [Yasuda 1962]. The latter means that when PET and water vapor are in thermal equilibrium, the water density in PET is a factor 11.1 higher than in the water vapor. For a relative humidity of 100%, this comes down to 0.0062 g of water per g of PET. Yasuda and Stannett also observed that at room temperature, Henry's law is obeyed for all partial pressures of water vapor and that in contact with liquid water, the same water uptake was reached as in contact with 100% saturated water vapor [Yasuda 1962]. Bellucci and Nicodemo did both capacitive and gravimetric diffusion measurements on free PET films and found D values between $2 \times 10^{-9} \text{ cm}^2/\text{s}$ and $10 \times 10^{-9} \text{ cm}^2/\text{s}$ and water weight fractions between 0.0054 and 0.0095 in contact with liquid water.

Another interesting observation is that the temperature dependence of the diffusion coefficient of water in PET obeys Arrhenius's law, or in other words, that D is proportional to a factor $\exp(-A/k_B T)$ where A is an activation energy [Yasuda 1962, Michaels 1963, Light 1982]. This implies the presence of potential barriers for the water transport, unless PET is

hydrophobic and the water molecules will not be bonded to the polymer chains. The potential barriers will probably be narrow tunnels in the free volume of PET.

When the crystalline fraction of PET increases, water transport becomes more difficult. Lasoski and Cobbs observed that both the solubility and the diffusion coefficient of water in PET decreases with the crystallinity [Lasoski 1959]. They found that the weight fraction of water decreases linearly from 0.0061 for amorphous PET to 0.0035 for 40% crystalline PET. The permeability rapidly decreases by about a factor 2 when the crystallinity increases from zero to 10-15% and eventually becomes a factor 3 lower when the crystallinity further increases to 40%.

Not only heat treatments but also deformation processes may influence the crystallinity of PET. This influence of deformation on the transport properties of PET is relevant for the application in PET bottles, because the PET is highly deformed when it is blown into a mould in order to create a bottle. For the packaging of carbonated beverages, the transport of oxygen and carbon dioxide is most important. Slee, Orchard, Cobbs and Bower found that for uniaxial draw ratios from 1 to 4 and for biaxial draw ratios from 1 to 3, the permeability of the gases in PET gradually decreases by factors 1.8 and 1.6 respectively [Slee 1989]. Also for our purposes, which is the application of PET coated ECCS for food and beverage cans, the influence of deformation on the transport properties of PET is important as mentioned in chapter 1. However we are rather interested in the transport of water and ions.

Much less is known about the transport of ions in PET than about the transport of water and gases. In fact, we only found an estimate for the upper limit of the diffusion coefficient of Na^+ and Cl^- ions in PET from an experiment of Kloppers, Bellucci, Latanision, and Brennan, who mounted a free PET film of 22 μm thickness between two chambers filled with NaCl solutions in water. They applied 10V over a PET film and measured the electric resistance of the film as a function of time. They eventually expected a sharp decrease of the resistance when the ion diffusion fronts would meet, but since this did not happen in almost one month, they only could give an upper limit of $1 \times 10^{-13} \text{ cm}^2/\text{s}$ for the diffusion coefficient [Kloppers 1993]. The actual diffusion coefficient of ions in PET is probably even much lower, because the authors applied an electric field to accelerate the ion transport.

2.4.2 Gravimetric diffusion measurements on PET films

We did gravimetric diffusion measurements on free, amorphous PET films of 25 μm thickness that were supplied by Corus RD&T. The PET base material was produced by Eastman. Round samples with a diameter of 5 mm were cut out of sheets of PET foil and first kept in air with a relative humidity of 100%. It was assumed that after one day, the samples were saturated. Then the samples were mounted in a microbalance, a dry air stream was applied to remove the water evaporating out of the samples and the mass of the samples was followed as a function of time. The measurements were carried out at room temperature.

A typical gravimetric measurement is plotted in figure 2.9. The measurements showed good agreement with equation (2.2) and with figure 2.6. This suggests that the diffusion of water out of the PET film obeys Fick's laws with only a single diffusion coefficient. This is in agreement with the hydrophobic behavior of PET, which only allows free volume transport.

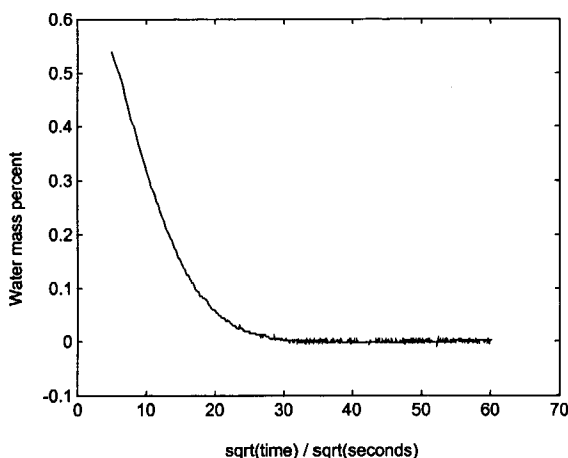


Figure 2.9. Gravimetric diffusion measurement showing the evaporation of water out of a free PET film of $25\mu\text{m}$ thickness. The measurement starts only after about 30 seconds, which is the time required for mounting the samples in the microbalance.

Measurements were done for a large number of samples and the diffusion coefficient was determined by fitting, as discussed in section 2.3.1. This led to an average value of $D = 4.0 \times 10^{-9} \text{cm}^2/\text{s}$ for the diffusion coefficient of water in the PET films. For the separate measurements, D varied between $3.6 \times 10^{-9} \text{cm}^2/\text{s}$ and $4.4 \times 10^{-9} \text{cm}^2/\text{s}$, which may be due to actual difference of D between the samples or due to the first 30 seconds in which the samples were mounted in the microbalance. Although this was done in dry air and with great care, the well-defined evaporation actually only starts when the dry air stream is switched on.

By extrapolation of the initial part of the curve to time zero, the total mass fraction of water in PET in air with a relative humidity of 100% could be estimated. This led to mass fractions between 0.003 and 0.008 of water in PET with an average value of 0.006. Considerably more accurate measurements of the solvability were done at Acordis R&D by measuring the mass of our PET films at 0%, 10%, 20% and so on up to 90% relative humidity and at room temperature (20°C). These measurements are shown in figure 2.10. A linear behavior was observed, which means that Henry's law was obeyed, in agreement with earlier observations [Yusada 1962]. By extrapolation, a mass fraction of 0.0078 of water at 100% relative humidity was determined. A value of $1.335 \text{g}/\text{cm}^3$ was determined for the density of our PET films, which corresponds to literature values for amorphous PET.

We also did some gravimetric measurements on supported PET films on ECCS. This however did not lead to reproducible measurements, which is probably due to corrosion of the bare steel on the edges of the samples. Sealing of the edges by some kind of kit was considered, but this introduces new systematic errors due to the water uptake of the kit.

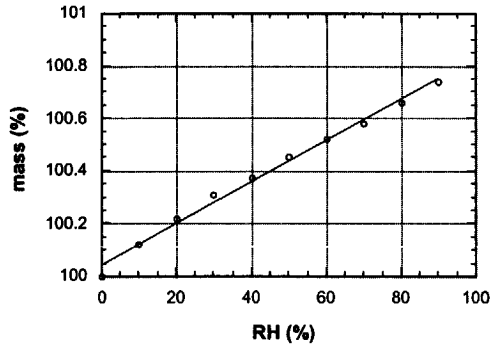


Figure 2.10. The mass of a free PET film as a function of the relative humidity (RH). The mass is normalized to the mass of the PET film in dry air.

2.4.3 Electric impedance measurements of water and ion uptake

The electric impedance measurements were done with the 2500A 1kHz Ultra-Precision Capacitance Bridge of Andeen Hagerling. This apparatus measures the real and imaginary parts of the impedance between two electrodes at a frequency of 1 kHz. The user can choose to automatically translate the measured impedance to a parallel capacitance C and conductance S . In this mode we followed C and S of amorphous PET films of $25\mu\text{m}$ on ECCS as a function of time after immersion by electrolyte. According to the manufacturer, the films were identical to the free films used for the gravimetric diffusion measurements. Prior to immersion, the samples were dried for at least 1 day at 50°C . It was assumed that after this pretreatment, the free volume water had evaporated.

Below we first discuss the measurements of the capacitance and then the measurements of the conductivity. The measurements never showed remarkable differences, even after rather harsh treatments of the PET coating, which led to suspicion on the method and the experimental setup. Therefore we did some additional verification tests, which are presented in the end of this section.

Capacitive uptake measurements

In figure 2.11, a typical capacitance measurement is shown of PET coated ECCS under immersion of an electrolyte. We see that plotted as a function of the square root of time, the capacitance C shows two straight asymptotes, similar to the gravimetric measurements. This not only suggests that a single D value can describe the water transport, but also that all water molecules have the same contribution to the capacitance. Both statements are in agreement with the expectation that water is only transported through the free volume of PET.

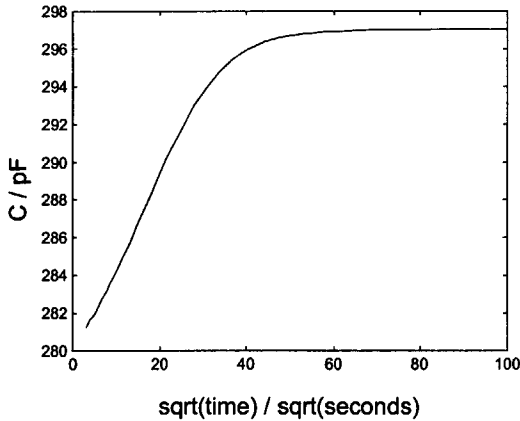


Figure 2.11. Typical capacitance measurements on a PET film on ECCS during water and ion uptake from an electrolyte.

We fitted the diffusion coefficient of water transport in PET by equation (2.3), with L replaced by $2L$, since in this experiment we were dealing with supported coatings. In addition we studied the influence of:

- Deformation of the samples;
- An initial heat treatment prior to the diffusion measurements;
- An alcohol content of 5% in the electrolyte;
- Different pH values and different acids.

All these influences are relevant to the application of PET coated ECCS for food and beverage cans, as discussed in chapter 1. The influence of the deformation was studied for uniaxially and biaxially stretched samples that were stretched up to 25% for the uniaxially stretched samples and up to 20% for the biaxially stretched samples [Jong, de 1997]. As discussed in section 2.4.1, it was expected that the diffusion rate would decrease with the deformation. Also fragments of cans after two deep drawing steps and after wall ironing with three rings were studied, but this mostly led to irreproducible measurements with large fluctuations of C and S as a function of time. This suggests that on the can fragment, the PET coatings were too much damaged for this type of measurement.

Four types of initial heat treatments were applied prior to water uptake measurements. These were 15 or 60 minutes of pasteurization at 75°C and 15 or 60 minutes of sterilization at 121°C and 2 bar. The heat treatments were done under immersion of buffered citric acid of $\text{pH} = 3$, because pasteurization and sterilization are done while the food or beverage product is inside the can. (A recipe for buffered citric acid with $\text{pH} = 3$ is given in chapter 3.)

The influence of the pH , of different acids and of a 5% alcohol content was studied to model different food and beverage products. We used phosphoric acid of $\text{pH} = 2.5$, buffered citric acid of $\text{pH} = 3$ and sulfuric acid of $\text{pH} = 1, 2, 3$ and 4. The influence of an alcohol content of 5% was studied because in contrast with water, PET is well wetted by alcohol and this might influence the transport properties. An alcohol content of 5% was studied because

this corresponds to most kinds of beer. In fact we have observed that although PET is hydrophobic, it is not “beer-phobic” and well wetted by beer.

All measurements showed similar capacitance plots as a function of time and no significant differences were observed that could be related to one of the treatments mentioned above. The diffusion constant was always between $3.0 \times 10^{-9} \text{cm}^2/\text{s}$ and $4.0 \times 10^{-9} \text{cm}^2/\text{s}$ and the relative increase in the capacitance of the coating was between 5 and 7% in saturation. The observed differences may well be due to the uncertainty in the thickness L of the coating (which varied between 20 and $30 \mu\text{m}$). This does no justice to the accuracy of the capacitance bridge, which measured the capacitance values of about 300pF with a precision of about 0.01pF.

The reason for the fact that we observed no significant differences as a function of deformation is probably that the deformation was not very large for the samples that were not too much damaged. As discussed in section 2.4.1, a reduction of the permeability by a factor 2 by deformation has been observed [Slee 1989], but for much higher deformations (a draw ratio of 4) and for oxygen instead of water.

That pasteurization (at 75°C) did not alter the transport properties of PET is not surprising, because crystallization of PET only starts between 110 and 130°C . That sterilization (at 121°C) had no effect is more remarkable. In fact after 1 hour of sterilization in buffered citric acid, the PET coatings had the typical opaque appearance of crystalline PET. They also were yellowish, which is probably due to the penetration of citric acid. After 1 hour at 120°C , the crystallinity of PET may be 35% [Michaels 1963] and this should reduce the diffusion rate of water in PET by a factor 2 [Lasoski 1959]. However, this was not observed and we have no explanation for this observation. Perhaps the PET of the base material of PET coated ECCS already contained a crystalline fraction of 10-15%. Above this fraction, the transport properties are not much further reduced [Lasoski 1959].

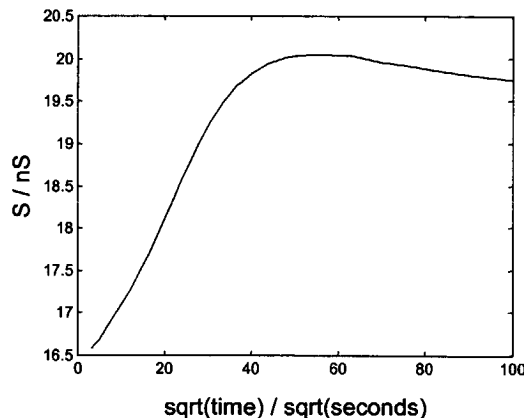


Figure 2.12. Typical conductance measurements on a PET film on ECCS during water and ion uptake from an electrolyte.

Water and ion uptake measurements with the conductance

In figure 2.12, a typical uptake curve for the conductance S is shown. The initial large increase in S occurs at the same time as the large increase in C and is quite certainly due to the water uptake, by which ϵ'' increases simultaneously with ϵ' . In figure 2.12, S reaches a maximum and then decreases to a stationary value. This decrease was not always measured. Equally often no decrease or a slight increase was measured. As discussed in section 2.3.2, we had expected an increase of S due to ion uptake.

The ratio of the conductance S and the capacitance C during uptake was always between about 0.060 and 0.067 nS/pF. Ignoring the conductivity γ this comes down to a ratio of ϵ''/ϵ' of about 0.010, which is in fair agreement with literature values on PET (cf. [Reddish 1950, Willems 1995]). Only after 1 hour of sterilization, a significant change of the ratio S/C was observed, which was then between 0.043 and 0.050 nS/pF. This shows that sterilization did affect the dielectric properties of the PET films, although it did not noticeably affect the water transport.

The observation that different acids did not change the results for S as a function of time also suggests that we only measured the water uptake and not the ion uptake. Apparently the ion uptake of our PET films is too slow for these measurements. In order to get some information on the ion transport in PET, we therefore did some additional long-term experiments, which are discussed in section 2.4.4.

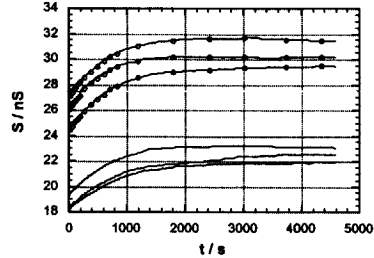
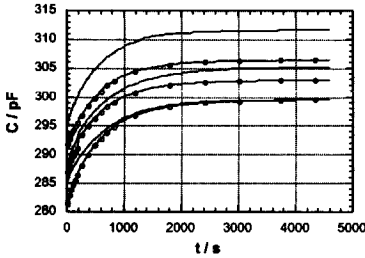
Verification of the technique

The measurements of the conductance S during uptake and the fact that no clear differences were found in the diffusion coefficient after heat and deformation treatments led to suspicion. Could it be that the observed increment of the capacitance was due to something else than water uptake, for instance due to the increase of the contact area between the electrolyte and the top of the coating? In that case, we would have measured nothing than the rate at which the electrolyte creeps under the cylinder containing the electrolyte. Then the breakpoint time is determined by the fixation of the cylinder on the coating and is indeed independent of treatments such as sterilization and so on. In fact, the breakpoint time would then even be independent of the coating thickness, leading to a diffusion coefficient that seems to increase with the coating thickness. Indeed this is sometimes reported in literature by authors using a clamp cell with a rubber ring (cf. [Bellucci 1993, Kloppers 1993]). Hence variation of the thickness of the coating is one way to verify whether one really measures water uptake or only surface enlargement. In our case however, variation of the coating thickness was not easily possible and therefore we did a number of other verification tests:

- Compare different ways of fixation of the cylinder to the coating;
- Use cylinders with different areas;
- Measure the uptake of a free PET film;
- Measure the uptake for different types of coatings.

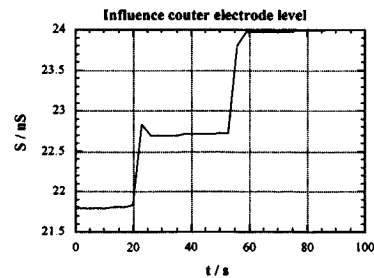
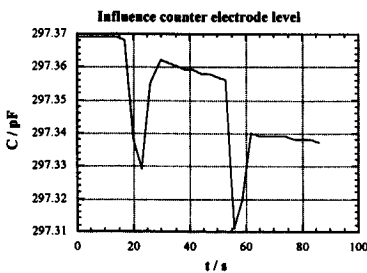
We examined three ways of fixing the cylinder on top of the PET coating, silicon kit, high vacuum grease and mechanically clamping via a rubber ring. The first two fixation methods lead to about the same uptake measurements for both the capacitance C as well as for the conductance S , but for the clamp cell, completely different results were obtained. The value

of C then was about the same, but the ratio S/C was about 0.3nS/pF, which is 5 times larger than for silicon kit or vacuum grease. When one and the same rubber ring was used for some time, the measurements reproduced rather well, but when a new or a dried ring was taken, spectacularly high increments of C and S were measured during water uptake. We therefore abandoned the clamp cell for water uptake measurements.

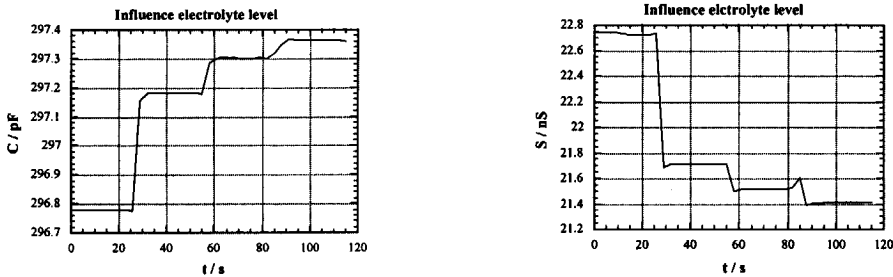


Figures 2.13 and 2.14. *The capacitance and conductance of supported PET coatings during water uptake for immersed areas with diameters of 2.0 and 4.4cm. The curves with markers are the measurements for the 4.4cm cylinder after scaling by a factor $(4.4/2.0)^2$.*

The cylinders containing the electrolyte were Perspex cylinders with an inner diameter of 2.0cm and an outer diameter of 3.0cm. We did also some water uptake measurements using a cylinder with an inner diameter of 4.4cm. In figures 2.13 and 2.14 these measurements are compared to some results for a 2.0cm diameter cylinder. Since both the capacitance C and the conductance S of the coating are proportional to the immersed area, C and S and also to the increase of C and S during water uptake should be scaled by a factor $(4.4/2.0)^2 = 4.84$. As follows from figures 2.13 and 2.14, this holds for C and for the increase in C during water uptake, but not for S and the increase in S, which scale by factors of about 6.6 and 5.4 respectively. This suggests that the measured C value is close to the capacitance of the immersed part of the PET coating, but that the measured S value also contains other contributions.



Figures 2.15 and 2.16. *Influence of the counter electrode level on the capacitance and the conductance. The counter electrode is stepwise lifted with time.*



Figures 2.17 and 2.18. Influence of the electrolyte level on the capacitance and the conductance. The electrolyte level is stepwise increased with time.

The additional contributions to S can originate from the environment via the Perspex cylinder and from the electrolyte resistance between the counter electrode and the top of the coating. We studied the influence of the electrolyte by two simple experiments:

- Increasing the counter electrode level from 0.5 to 2.0cm while the electrolyte level was 2.0cm. Then C decreased by 0.01%, while S increased by 10%;
- Increasing the electrolyte level from 0.5 to 2.0cm with the counter electrode at 0.5cm. Then C increased by 0.2% while S decreased by 6%.

These measurements are shown in figures 2.15 to 2.19. By elementary network analysis one can show that when the electrolyte resistance increases, the real part of $1/Z$, which is the apparent S, increases while C decreases. This agrees with both measurements, since in the first experiment the electrolyte resistance is increased while in the second experiment, the electrolyte resistance is decreased. Also for the second experiment, the coupling between the counter electrode and the environment is increased. It follows that for the measurement of C, the counter electrode and electrolyte levels are not so critical, but for the measurement of S should be kept as low as possible. The minimum distance is obtained when a metal layer is sputtered on top of the polymer and used as a counter electrode, as for permittivity measurements on dry polymers. For water uptake measurements this is no obvious solution, since such a metal layer will block the water transport from the electrolyte to the coating.

We also measured the water uptake of a free PET film by electric impedance measurements. Following the discussion of sections 2.3.1 and 2.3.2, this should lead to a 4 times smaller breakpoint time, which provides another test of the technique. In fact, the breakpoint time of a free PET film was already measured by our gravimetric measurements. This led to a value of $D = 4.0 \times 10^{-9} \text{ cm}^2/\text{s}$, which is in good agreement with our capacitive measurements on supported PET films. Nevertheless we also did some electric impedance measurements on free PET films by mounting a free PET film between two Perspex cylinders. The result is presented in figure 2.19, in which also a measurement on a supported PET film is plotted. From this figure it follows that the breakpoint time is indeed about 4 times smaller for the free PET film. In addition we see that the total increase of C is about the same for the free

and supported PET films. These observations also suggest that the transport properties of the free and supported PET films are about the same, as claimed by Corus RD&T.

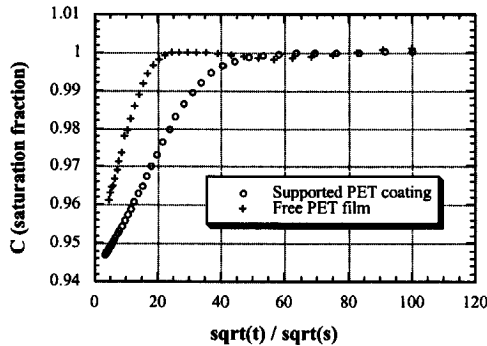


Figure 2.19. *Water uptake of a free film compared to a supported film. The square root of the breakpoint time is about two times larger for the supported film. The total relative increase of the capacity is about the same.*

It is interesting to compare the total increase of the capacitance to the total mass increase for free PET films, because this allows us to derive the actual influence of a given water mass fraction on the real part of the permittivity ϵ' . For the free films, we found a total increment of 6.0% of the capacitance, which is about the same as for the PET coatings on ECCS. The total mass increment was about 0.78%, corresponding to a volume fraction of liquid water of about 1.04%. We can compare these values via the Brasher-Kingsbury mixture formula:

$$\ln \epsilon' = (1 - v_w) \ln \epsilon'_{\text{PET}} + v_w \ln \epsilon'_w$$

This can be rewritten into the following expression for the volume fraction of water:

$$v_w = \frac{\ln(\epsilon' / \epsilon'_{\text{PET}})}{\ln(\epsilon'_w / \epsilon'_{\text{PET}})} = \frac{\ln(C / C_{\text{PET}})}{\ln(\epsilon'_w / \epsilon'_{\text{PET}})}$$

With $\epsilon'_w = 80$ and $\epsilon'_{\text{PET}} = 3.8$ [Willems 1995] we find that an increase of 6% of the capacity corresponds to a volume fraction of 1.9% of liquid water in PET. This is significantly larger than the volume fraction of 1.04% found by our gravimetric measurements. This difference cannot be explained by the uncertainty in ϵ'_w since this would require a value of $\epsilon'_w = 5500$, while ϵ'_w should lie between 1 and 80.

A final check for verifying the measurement technique was to measure the water uptake of a completely different type of coating. We chose an epoxy coating on aluminum and the measurement is shown in figure 2.20. We see that for this epoxy coating, which was considerably thicker than our PET coatings (about 80 μm versus 25 μm), the water uptake

goes much slower. In case the increase of the capacitance would have been only due to surface enlargement, the same breakpoint time would have been measured as for the PET coatings, but this is clearly not the case. In addition, the capacitance further increases after a kink, which is characteristic for the slow uptake of bonded water in hydrophilic coatings (cf. [Westing, van 1994]). Taking the time at the kink as the breakpoint time, an diffusion coefficient of about $4 \times 10^{-9} \text{cm}^2/\text{s}$ follows. This is probably the diffusion coefficient of the initial free volume transport of water in the epoxy coating.

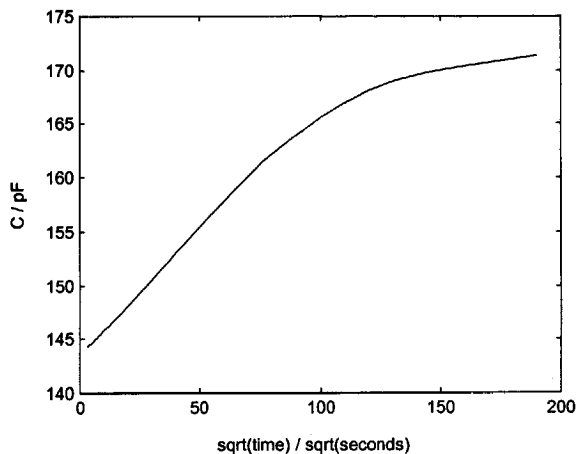


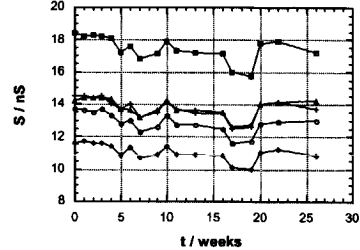
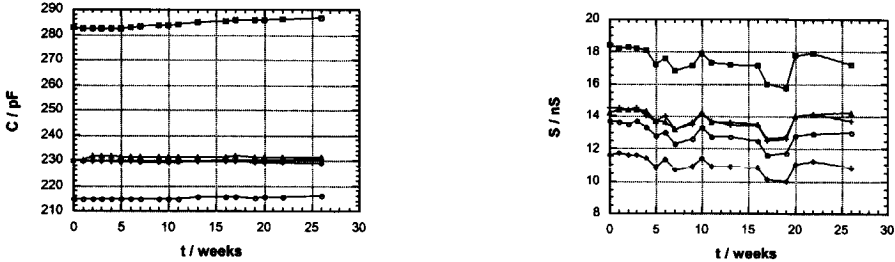
Figure 2.20. Water uptake of an epoxy film of about $80 \mu\text{m}$. The uptake in this hydrophilic coating is much slower than in PET. Also there is no constant plateau value, but a further increase with a different diffusion rate.

2.4.4 Long-term measurements on water and ion transport in PET

With the techniques of section 2.4.2 and 2.4.3, we have in fact only studied the water transport in PET for relatively short times. The water transport for longer times such as a few months is also very important, because it is often due to the slower penetration of water in the polymer structure that the coating finally breaks down and fails to protect the underlying metal. Further we had not observed any effect of the ion transport in PET and therefore we also followed the electric impedance of a number of PET coatings on ECCS during a considerably longer time, namely for half a year. The electrolyte was in this case phosphoric acid with a pH value of 2.5. The results are plotted in figures 2.21 and 2.22. We see that neither the capacitance C nor the conductance S changed significantly. The fluctuations of about 0.1% in the capacitance and of about 5% in the conductance are probably due to temperature fluctuations, since they overlap for the different samples.

No breakdown of the PET coating was therefore observed within half a year, in contrast with earlier observations on many other types of corrosion protective coatings [Wormwell 1949, Westing, van 1994]. This could be due to the hydrophobic behavior of

PET. In contrast with hydrophilic polymers such as epoxy's, the water will only fill the free volume and not penetrate in the polymer structure itself, which remains intact.



Figures 2.21 and 2.22. Capacity and conductance of a PET coating on ECCS during half a year of immersion in phosphoric acid of pH = 2.5.

By visual inspection, we observed no corrosion of the ECCS underneath the PET within 6 months, which may be due to the slow ion transport through the PET, but also to the good adhesion of the PET coating on the ECCS. To verify this, we did a similar experiment with a free PET film of the same thickness on a polished iron surface. On the PET film we mounted a cylinder which we filled with sulfuric acid of pH = 1. We closed the cylinder to prevent oxygen to enter the electrolyte. In half a year the iron finally got humid, but by visual inspection we observed no corrosion. Hence, the PET film effectively protected the iron from corrosion, which means that the H^+ ions were not significantly transported through the PET during half a year. Note that with this experiment, one is essentially measuring the transport of H^+ and SO_4^{2-} together, since diffusion processes transport no net electric charge. From this experiment, a lower limit of half a year or 1.6×10^7 s can be given for the breakpoint time of H^+ and SO_4^{2-} , from which an upper limit of $D = 3 \times 10^{-13} \text{ cm}^2/\text{s}$ can be derived for the diffusion coefficient. This value agrees with the upper limit of $D = 1 \times 10^{-13} \text{ cm}^2/\text{s}$ for the transport of Na^+ and Cl^- through PET from Kloppers, Bellucci, Latinis and Brennan [Kloppers 1993]. The comparison is however not completely fair because they used different ions and accelerated the ion transport by a direct voltage of 10V.

A final indication that the ion transport in our PET films is extremely slow is given by a charge decay measurement. For this experiment, we charged the top of a PET coating on ECCS by rubbing with a tissue and we observed with the Kelvin Probe that it took about one day before half of the charge had flown away. (In chapter 6, the measurement of charges on top of coatings by the Kelvin Probe method is described.) Assuming that the charge had flown away through the PET, the time constant of the charge decay is equal to $C/S = \epsilon'/\gamma = 10^5$ s (one day). Then with $\epsilon' = 3.8 \times 8.9 \times 10^{-12} \text{ F/m}$ [Willems 1995], we find a conductivity of $\gamma = 2 \times 10^{-16} \text{ S/m}$ for our PET films. This value means for instance that the resistance of our 25 μm PET films is equal to 500T Ω for an area a diameter of 2cm. This illustrates how small the conductivity value is. By the way our conductivity value is much smaller than the value of $4.3 \times 10^{-8} \text{ S/m}$ reported by Willems, who determined the conductivity of PET by the

Kramers-Kronig transform [Willems 1995]. These values are not necessarily conflicting, because the ion content of our PET coatings may indeed have been much smaller.

2.5 Summary and discussion

We followed the water transport through PET films by gravimetric and electric impedance measurements. For the gravimetric measurements, we only studied free PET films and for the impedance measurements we both studied free films as well as supported PET films that were PET coatings on ECCS. All PET films had a thickness of 25 μm . The gravimetric and impedance measurements showed good agreement and both could be very well described by Fick's laws with a single diffusion coefficient. This suggests that the PET coatings were homogeneous and that the water molecules were only transported by one and the same process, which is probably free volume diffusion. From the capacitive measurements it further follows that all water molecules have the same contribution to the capacitance of the coating. This means that they are probably all in similar positions, which will be not bonded and in the free volume of the PET structure.

All measurements gave a value of $D = 4 \times 10^{-9} \text{ cm}^2/\text{s}$, with a variation of about 15% between the different samples. These variations were probably due to the local variations of 10% of the thickness of the PET films. (This does not mean that D depends on the coating thickness, but since D is estimated from the breakpoint time which is proportional to the square of the coating thickness, the estimator for D also depends on the coating thickness.) For the solubility of water in PET we found a value of 0.0078g of water per g of PET from the gravimetric measurements for free PET films in air with a relative humidity of 100%. Both the values of the diffusion coefficient and the solubility are in good agreement with literature values. We also determined the solubility from the increment of the capacitance during water uptake by using the Brasher-Kingsbury formula. The 6% increment of the capacity implies a volume fraction of 1.9% of liquid water and corresponds to a mass fraction of 0.014 of water in PET. This value is almost a factor 2 larger than the value of 0.0078 from the gravimetric measurements.

The disagreement in the results for the solubility as determined from the capacitive and the gravimetric measurements is probably due to the invalidity of the Brasher-Kingsbury relation. In fact gravimetric water uptake measurements can also be inaccurate due to water sticking on the surface. But in our case, where we determined the solubility by extrapolation of the mass of PET films in a relative humidity of 0%, 10%, 20% and so on, this error was probably not made. It is a challenge to find a better model of the influence of the free volume water on the permittivity of PET. Perhaps computer simulations of the average permittivity of characteristic unit cells of water in PET can help to clarify this problem.

It was expected that heat treatments or deformation would affect the transport properties of PET, because these treatments or influence the degree of crystallinity of PET. We studied the influence deformation by uniaxially and biaxially stretched samples and by fragments prefinished cans. Four different heat treatments were studied, namely 15 or 60 minutes of pasteurization at 75°C and 15 or 60 minutes of sterilization at 121°C and 2 bar. After all these treatments however, the same values of the diffusion coefficient and the solubility were found as for untreated PET films. For pasteurization this is not very surprising

because crystallization of PET only starts between 110 and 130°C. Also for the lower degrees of deformation (draw ratios of up to about 1.3), no large effect was expected since from literature it was known that only at draw ratios of 4 the permeability of PET is increased by a factor 2. That sterilization had no effect is more surprising, since the sterilization temperature lies well in the crystallization region. In fact after 1 hour of sterilization, a crystalline fraction of 35% was expected and indeed after this treatment, our PET coatings had the typical opaque appearance of (partly) crystalline PET. However, the transport properties were not affected. Perhaps this is explained by the presence of a crystalline fraction in the untreated samples as delivered by Corus RD&T.

We also studied the effect of some different additives to the electrolyte. In case the additive product interacts with PET, it could break down the polymer structure of the coating and alter its transport properties. In fact for hydrophilic coatings such as epoxy lacquers, water itself can have a destructive effect on the polymer structure of the coating, but for hydrophobic polymers such as PET, this is not expected. We have studied the water uptake in H₂SO₄ of pH values between 1 and 4, in H₃PO₄ of pH = 2.5 and in buffered citric acid of pH = 3. Also we studied the influence of a 5% ethanol content as in beer, since PET is well wetted by ethanol. In none of these cases we observed a significant difference in the rate of water transport, but other food and beverage additives may interact more strongly with PET.

The rapid water uptake measurements mentioned above gave no clue about the long term effect of immersion in water based electrolytes and on the ion transport in PET. Therefore we also did a number of measurements during half a year, which is in the order of the shelf life of a number of canned products. These measurements were done in phosphoric and sulfuric acid. We found that during half a year, water does not break down the polymer structure of PET and that the ion transport was negligible. An upper limit of $3 \times 10^{-9} \text{ cm}^2/\text{s}$ for the diffusion coefficient of H⁺ and SO₄²⁻ could be extracted from these observations. In addition we found that a good adhesion of the PET coating on the underlying ECCS is not necessary for preventing corrosion in the mentioned anorganic acids. Whether these conclusions also hold for other food and beverage additives is subject for further study.

Chapter 3

EIS on artificial defects in PET coated ECCS

3.1 Introduction

Under atmospheric conditions, oxygen and water may cause corrosion. And since polymers are generally no good blockers for water and oxygen transport, the corrosion protection of polymer coatings in atmospheric conditions will normally rely on a good adhesion between the coating and the underlying metal. In food and beverage cans however this is different. Water is available but oxygen is largely removed (to prevent oxidation of the stored product), and water and hydrogen ions will mainly cause corrosion. Now since polymers are good blockers for ions, the corrosion protection of a polymer coating inside consumable cans may be due to ion blocking as well as a good adhesion of the coating with the can metal.

The transport of water and ions through PET coatings was already studied in the previous chapter and we found that the ion transport in PET is probably low enough to prevent corrosion during shelf life. In addition we found that PET does not break down by the action of water. Hence a good adhesion to the underlying metal might not even be necessary for the corrosion protection of PET in food and beverage cans. However, there are indications that some particular contents of consumables may gradually break down PET. Then the adhesion of PET to the underlying metal becomes important to postpone corrosion. For this reason we also studied the adhesion of PET on ECCS. We did this by electrochemical impedance measurements, as discussed below.

3.2 Electric impedance spectroscopy

The concept of impedance was invented by Heaviside between 1880 and 1887 for rapidly finding the stationary solutions of systems of linear differential equations. Heaviside initially applied his technique to Maxwell's equations, which he had studied and simplified considerably. His solution method was based on the addition of an imaginary part to sinusoidal solutions and to work with complex exponentials. Heaviside's method surprised

the mathematical world, until the equivalence with Fourier transforms was pointed out by Bromwich. It is the latter treatment that will be followed in the discussion of this section.

3.2.1 Systems and signals

With electric impedance measurements, the electric properties of a system between two electrodes are studied. This is generally done by applying an alternating voltage over the electrodes and by measuring the resulting current through the medium, but the opposite (applying a current and measuring the voltage) is also possible and sometimes done. The procedure may be represented by the following scheme:

$$\text{input } x(t) \Rightarrow \boxed{\text{system}} \Rightarrow \text{output } y(t)$$

The input signal $x(t)$ can thus either be an alternating voltage or current and for the output signal $y(t)$ it is then the opposite. This scheme in fact describes any type of experiments in which one determines certain properties of a system by studying the relation between input and output signals $x(t)$ and $y(t)$. Assuming that the output is determined by the input, one can seek for a relation between the output and the input characterizing the system. As we will discuss, such a relation will generally be a nonlinear differential equation containing the input and output signals. Under some conditions however the system may even be characterized in a more compact way by a “system function”. The electric impedance is an example of such a system function, characterizing the electric properties of a system. In this section we will give a short discussion on the characterization of systems leading to the fundamentals of the electric impedance. For a more extensive treatment, the reader is referred to the book of Oppenheim, Willsky and Young [Oppenheim 1983].

It will be assumed that the signals $x(t)$ and $y(t)$ are “well-behaved” functions of time. Technically this means that around any particular time t_0 , they can be represented by a Taylor-expansion. Then at times t around t_0 , the signals are completely characterized by the function values $x(t_0)$ and $y(t_0)$ and the series of first, second and higher order derivatives $x'(t_0)$, $y'(t_0)$, $x''(t_0)$, $y''(t_0)$ and so on, together with the time interval $dt = (t - t_0)$ from the time t_0 . The most general relation between $x(t)$ and $y(t)$ is then a relation containing these quantities and is a nonlinear differential equation. (In case the signals are not well-behaved functions such as fractals or noise, fractional derivatives may be necessary to describe the relationship between $x(t)$ and $y(t)$.) A special group of systems is the group of linear time invariant systems (also abbreviated as LTI systems). Such systems have the following properties:

- They are time invariant, meaning that a time shift in the input leads to the same time shift in the output. Hence, a system reacts the same on the same input signal at any time;
- They are linear. This means that if two input signals $x_1(t)$ and $x_2(t)$ lead to outputs $y_1(t)$ and $y_2(t)$, then the input $\alpha x_1(t) + \beta x_2(t)$ leads to the output $\alpha y_1(t) + \beta y_2(t)$.

LTI systems are described by linear and time invariant differential equations. These have the following general form:

$$\sum_n a_n \frac{d^n y}{dt^n} = \sum_m b_m \frac{d^m x}{dt^m} \quad (3.1)$$

In this expression, the indices m and n run from zero to infinity although in practical models, the coefficients a_n and b_m are often zero above certain m and n values. (Note that the “zeroth order derivative” of a function is the function itself.) The coefficients a_n and b_m characterize the system and are independent of time when the system is time invariant. More generally (for nonlinear differential equations), the time must not appear explicitly in the system equation for a time invariant system. The linearity of the system on the other hand basically relies on the absence of nonlinear forms such as y^2 and $\sin(dx/dt)$ in the system equation.

LTI systems are relatively easy and convenient to study, in particular for reasons of the uniqueness of the solution and for the characterization by system functions, as will be discussed. Therefore LTI systems are used for physical models in many different research areas such as mechanics, electromagnetism and quantum mechanics. (In fact, the tendency to describe phenomena by LTI systems is so strong that essentially nonlinear phenomena such as chaos have only become popular rather recently.)

It is rather safe to say that in practice, no system will be actually linear and time invariant. However, under some well-defined conditions many systems do behave as LTI systems in good approximation. These conditions are:

- The variations of the system with time are relatively slow with respect to the measurement time. The system can then be considered time invariant;
- The input and output signals are only small variations around a certain setpoint. The relation between the input and output signals will then often be linear.

The first argument will need no further explanation. The second, indecisive argument can be illustrated by the following example of a system equation:

$$y = x^2$$

Note that in this equation, the derivatives of $x(t)$ and $y(t)$ are absent and the system equation is a functional relation. This system therefore responds instantaneously, hence with no time delay. In fact, for slowly varying signals, every system will respond in such a way. Now let us consider small variations around the setpoint $(x_0, y_0) = (2, 4)$. By writing $x(t) = 2 + \delta x(t)$ and $y(t) = y_0 + \delta y(t)$ we get in first order:

$$4 + \delta y = (2 + \delta x)^2 \cong 4 + 2\delta x$$

The setpoint solution $(x_0, y_0) = (2, 4)$ cancels and we arrive at the following relation between the signals $\delta x(t)$ and $\delta y(t)$:

$$\delta y \cong 2\delta x$$

which is a linear relation. It is easily shown that small perturbations δx and δy around any other setpoint (x_0, y_0) also leads to a linear system, except for $(x_0, y_0) = (0, 0)$. Then:

$$\delta y = \delta x^2$$

and this is a nonlinear relation, even for arbitrarily small perturbations. We see that it is not generally true that the relation between small perturbations δx and δy is described by an LTI system, although it will often be the case.

Now we turn our attention to the characterization of LTI systems by a system function. In fact, the coefficients a_n and b_m of the system equation are a complete characterization of the system, but for techniques such as impedance measurements where the system is studied by the response to sinusoidal input signals, this is not the most suited description. In that case one prefers to work with a system function $H(\omega)$ giving the response to different frequencies. The natural way to arrive at such a system function is to work with the Fourier transforms of the signals. The Fourier transform and its inverse are defined by:

$$X(\omega) = \int_{t=-\infty}^{\infty} x(t)e^{-i\omega t} dt \quad (3.2a)$$

$$x(t) = \frac{1}{2\pi} \int_{\omega=-\infty}^{\infty} X(\omega)e^{i\omega t} d\omega \quad (3.2b)$$

In these equations, ω is the angular frequency, which is equal to $2\pi f$, where f is the frequency. The lower equation (3.2b), which is called the inverse Fourier transform, is in fact the expansion of a signal $x(t)$ on eigenfunctions $\exp(i\omega t)$. Taking the Fourier transform of equation (3.1) of an LTI system leads to the following:

$$\sum_n a_n (i\omega)^n Y(\omega) = \sum_m b_m (i\omega)^m X(\omega)$$

Here, the property is used that differentiation to the time t corresponds to multiplication of the Fourier transform by a factor $i\omega$ (this property immediately follows from the differentiation of equation (3.2b) with respect to the time). By taking $X(\omega)$ and $Y(\omega)$ out of the summation, one arrives at an expression of the following form:

$$Y(\omega) = H(\omega)X(\omega) \quad (3.3)$$

where the function $H(\omega)$ reads:

$$H(\omega) = \frac{\sum_m b_m (i\omega)^m}{\sum_n a_n (i\omega)^n}$$

Hence in terms of the Fourier transforms $X(\omega)$ and $Y(\omega)$, the differential equation of an LTI system comes down to the product $Y(\omega) = H(\omega)X(\omega)$, where $H(\omega)$ is the system function representing the system. Note that $H(\omega)$ is completely determined by the coefficients a_n and

b_m of the system equation (3.1). In case the coefficients a_n and b_m are zero above certain values of m and n , $H(\omega)$ is a polynomial fraction.

As the coefficients a_n and b_m , the system function $H(\omega)$ is a complete representation of the LTI system, giving the response to any given frequency ω . In particular when the input contains only one particular frequency, hence when $x(t) = A\exp(i\omega_0 t)$, we have $X(\omega) = A\delta(\omega - \omega_0)$ and $Y(\omega) = AH(\omega_0)\delta(\omega - \omega_0)$ and $y(t) = H(\omega_0)A\exp(i\omega_0 t)$. This means that when the input of an LTI system contains only one particular frequency ω_0 , the output will also only contains that frequency ω_0 . This is a very important characteristic of LTI systems, following immediately from equation (3.3).

It could be added that for other experimental techniques such as searching for oil in the earth crust, searching for cavities in pyramids and determining time constant distributions from dielectric spectra (see chapter 7), the frequency response function is not the most suited. Then a response function in the time or spatial domain is more appropriate, leading to pulse response functions, Green's functions or kernels of integral equations. These are different names used in different research areas for always the same idea of describing linear systems (cf. [Courant 1989]).

3.2.2 Real signals and systems

In the previous section we assumed that systems could be activated by a complex exponential input $x(t) = A\exp(i\omega_0 t)$. In reality however, one has to deal with real signals such as $\cos(\omega_0 t)$ or $\sin(\omega_0 t)$. However, these functions are not so far from complex exponentials since:

$$\begin{aligned}\cos(\omega_0 t) &= \frac{1}{2} \exp(i\omega_0 t) + \frac{1}{2} \exp(-i\omega_0 t) \\ \sin(\omega_0 t) &= \frac{1}{2i} \exp(i\omega_0 t) - \frac{1}{2i} \exp(-i\omega_0 t)\end{aligned}$$

Hence when a system is fed by a sine or cosine, it receives in fact the sum of two complex exponentials of frequencies $+\omega_0$ and $-\omega_0$ at one and the same time. We will now discuss the consequences of the "reality" of the input and output signals $x(t)$ and $y(t)$ and we will see that this leads to the elimination of negative frequencies from the theory, which is very pleasant since negative frequencies are unphysical. In terms of the Fourier transforms, the "reality" of the signals $x(t)$ and $y(t)$ comes down to the following relation:

$$\begin{aligned}X(-\omega) &= X^*(\omega) \\ Y(-\omega) &= Y^*(\omega)\end{aligned}$$

where the asterix denotes complex conjugation. This follows immediately from the fact that in expression (3.2a) for the Fourier transform, complex conjugation and replacing ω by $-\omega$ has the same effect. The same holds for the system function $H(\omega)$, since this is the ratio of $Y(\omega)$ and $X(\omega)$, hence one can also write:

$$H(-\omega) = H^*(\omega) \tag{3.4}$$

It follows that the reality of $x(t)$ and $y(t)$ implies that the function values of $X(\omega)$, $Y(\omega)$ and $H(\omega)$ for all negative frequencies are completely determined by the values for the corresponding positive frequencies. This allows one to entirely eliminate the negative frequencies from the theory, as will be discussed below. First we write the system function in polar coordinates by:

$$H(\omega) = |H(\omega)| \exp[i\phi_H(\omega)]$$

where $|H(\omega)|$ and $\phi_H(\omega)$ are the modulus and the phase of the system function. In terms of the modulus and the phase, property (3.4) implies:

$$|H(-\omega)| = |H(\omega)| \quad (3.5a)$$

$$\phi_H(-\omega) = -\phi_H(\omega) \quad (3.5b)$$

Hence the modulus is the same for negative frequencies and the phase is the opposite. These relations follow immediately from the fact that complex conjugation is the same as reflection in the real axis. Now let us determine the response to a real cosine input:

$$x(t) = A \cos(\omega_0 t) = \frac{1}{2} A \exp(-i\omega_0 t) + \frac{1}{2} A \exp(i\omega_0 t)$$

The Fourier transform of this input reads:

$$X(\omega) = \frac{1}{2} A \delta(\omega + \omega_0) + \frac{1}{2} A \delta(\omega - \omega_0)$$

The Fourier transform of the output signal is obtained by multiplication of $X(\omega)$ by the system function $H(\omega)$, giving:

$$Y(\omega) = H(\omega)X(\omega) = \frac{1}{2} AH(-\omega_0)\delta(\omega + \omega_0) + \frac{1}{2} AH(\omega_0)\delta(\omega - \omega_0)$$

where that $H(-\omega_0)$ and $H(\omega_0)$ could be substituted in the latter expression since these are the only relevant values in the product with the Dirac delta functions $\delta(\omega + \omega_0)$ and $\delta(\omega - \omega_0)$. Now by writing the system function in polar coordinates and using the above relations (3.5a) and (3.5b) gives:

$$Y(\omega) = \frac{1}{2} A |H(\omega_0)| [\delta(\omega + \omega_0) \exp[-i\phi_H(\omega_0)] + \delta(\omega - \omega_0) \exp[i\phi_H(\omega_0)]]$$

This finally leads to the following output signal:

$$y(t) = A |H(\omega_0)| \cos[\omega_0 t + \phi_H(\omega_0)]$$

This is a cosine with an amplitude that is a factor $|H(\omega_0)|$ larger and with a phase shift $\phi_H(\omega_0)$ with respect to input signal. This means that both the modulus and the phase of $H(\omega)$ can be determined by feeding the system with a cosine of frequency ω and then measuring the amplitude and the phase of the output signal. Hence the complete system function can be

obtained in this way with only real signals and without negative frequencies. This procedure is in fact essentially followed with electric impedance spectroscopy (EIS).

Before defining the electric impedance and discussing a number of relevant electric and electrochemical systems in the following section, we end this section with a few remarks on what happens when the basic assumptions on linearity and time invariance are violated. First when the system is varying with time, $H(\omega)$ is not a complete representation of the system. However when the variations with time are slow relative to the duration of the measurements, $H(\omega)$ will be constant during a measurement and the slow variations of the system with time can in fact be characterized by the variations of $H(\omega)$ with time.

On the other hand when the system is nonlinear, the response to a cosine of a certain frequency ω_0 will generally not be a cosine of that frequency. However when the system is stable and time invariant, a cosine input of frequency ω_0 will nevertheless lead to a stationary, periodic output signal with the same period $2\pi/\omega_0$ as the input cosine. This implies that the output can be written as a Fourier series:

$$y(t) = \sum_k B_k \cos[k\omega_0 t + \phi_k]$$

where $k = 0, 1, 2, 3$ and so on. In other words, $y(t)$ may also contain contributions of the harmonics of ω_0 . The amplitude and phase of only the ω_0 component are then an incomplete characterization of the system and instead of $H(\omega_0)$ one must also give the amplitude $B_k(\omega_0)$ and phase $\phi_k(\omega_0)$ of the response of the other harmonics of ω_0 . This may be illustrated by reconsidering the simple quadratic system:

$$y = x^2$$

Now in case an input $x(t) = x_0 + A \cos(\omega_0 t)$ is fed to the system, we have:

$$y(t) = [x_0 + A \cos(\omega_0 t)]^2 = x_0^2 + 2x_0 A \cos(\omega_0 t) + A^2 \cos^2(\omega_0 t)$$

By writing $y(t) = y_0 + \delta y(t)$, elimination of the setpoint $y_0 = x_0^2$ and rewriting the square of the cosine by elementary trigonometry, we arrive at:

$$\delta y(t) = \frac{1}{2} A^2 + 2x_0 A \cos(\omega_0 t) + \frac{1}{2} A^2 \cos(2\omega_0 t)$$

Clearly this is a nonlinear response, containing contributions of the zeroth harmonic (this is the constant term) and of the second harmonic. It is interesting to reconsider what happens when the amplitude A of the input becomes very small. Then the nonlinear terms (which are proportional to A^2) become negligible with respect to the linear term (which is proportional to A) and we arrive at a linear system, except for when we are at the setpoint $(x_0, y_0) = (0, 0)$ for which the linear term is always zero. Also note that none of the harmonics is shifted in phase with respect to the input. This is in agreement with the absence of derivatives in the system equation $y = x^2$, describing an instantaneously responding system.

It is often stated that, besides by checking for the presence of higher harmonics, the linearity of a system can also be tested by verifying whether the amplitude of the output is proportional to the amplitude of the input (cf. [Macdonald 1987]). The latter test relies on the basic property of linear systems that when an input $x(t)$ leads to an output $y(t)$, the input $\alpha x(t)$ leads to $\alpha y(t)$ for any α . This linearity test is in theory also sufficient, but may in practice not always be decisive. This is the case when only the amplitude of the first harmonics is measured, such as with a lock in amplifier (LIA). The reason is that for some nonlinear systems, such as the above system $y = x^2$, the amplitude of the first harmonics is proportional and linear to the amplitude A of the input. However, this is a very special example. For the system $y = x^3$ for instance, one can write the following for the response to the input $x(t) = x_0 + A\cos(\omega_0 t)$:

$$\delta y(t) = \frac{3}{2} x_0 A^2 + [3x_0^2 A + \frac{3}{4} A^3] \cos(\omega_0 t) + \frac{3}{2} x_0 A^2 \cos(2\omega_0 t) + \frac{1}{4} A^3 \cos(3\omega_0 t)$$

In this case, the amplitude of first harmonic is not proportional to A but also contains an A^3 contribution. Then the proportionality of the first harmonic contribution of the output (as measured by a LIA) with A is indeed sufficient for testing the linearity of the system. The will also hold for most other nonlinear systems.

3.2.3 Electric impedance spectroscopy

The above discussion may be summarized as follows. A system behaves as an LTI system under the following requirements:

- The system is time invariant with respect to the measurement time;
- The input and output signals are small perturbations around the setpoint;
- The lowest order response around the setpoint is linear.

Further we have seen that LTI systems can be characterized by a complex system function $H(\omega) = |H(\omega)|\exp(i\phi_H)$. This function can be determined for any ω by feeding the system with an input $x(t) = A\cos(\omega t)$ and by measuring the output that will be of the form $y(t) = B\cos(\omega t + \phi)$. The modulus of the system function $|H(\omega)|$ is then equal to the ratio of the amplitudes B/A of the output and the input and the phase ϕ_H of the system function is equal to the phase shift ϕ between the input and output signals. Finally, the linearity of the system may be tested by verifying whether the output contains higher harmonics of the input signal or by verifying whether the amplitude of the output signal is proportional to the amplitude of the input signal. All these issues applies to electric impedance spectroscopy (EIS).

As mentioned in section 3.2.1, electric impedance measurements are used to study the electric properties of a system mounted between two electrodes. We also mentioned two different implementations. In the first case, an alternating voltage δV is applied between the electrodes and the current response δI through the system is measured while in the other case it is the opposite. Assuming that the system is time invariant and linear in first order around a certain setpoint (V_0, I_0) , we have in the first case:

$$\delta I(\omega) = Y(\omega)\delta V(\omega) \quad (3.6a)$$

Here the system function $Y(\omega)$ is called the “admittance”. (The reader will not be confused with the Fourier transform of the output signal $y(t)$ of the previous sections.) In the second case the current is the input and the voltage is the output and we write:

$$\delta V(\omega) = Z(\omega)\delta I(\omega) \quad (3.6b)$$

Now in this case, the system function $Z(\omega)$ is called the “impedance”. It follows that the impedance and admittance of an LTI system are each inverse, or $Z(\omega) = 1/Y(\omega)$ and it is a matter of taste whether one describes the electric properties by the admittance or by the impedance. In literature, the impedance $Z(\omega)$ is more often used and also in this thesis we will basically work with $Z(\omega)$.

It could be mentioned that instead of applying a well-defined perturbation of known frequency, the impedance is also sometimes determined by noise analysis. The input and output signals are then the “random” noise from the environment. It is often suggested that the advantage of noise analysis is that the system is not disturbed, but this is ridiculous since one always disturbs a system when measuring, also by noise from the environment. And the perturbation by applying a cosine can be arbitrarily small by choosing the right amplitude. The only actual advantage of noise analysis seems to be the cheaper equipment (cf. [Mansfeld 1997]), but since the obtained $Z(\omega)$ is incomparably less accurate this will rarely be a serious consideration. We now proceed to a number of short discussions on the impedances of several network elements that are most significant for this work.

The impedance of resistors and capacitors

For this work, the two most important systems are the resistor and the capacitor. The relation between the current and the voltage over a resistor reads:

$$I(t) = \frac{1}{R} V(t)$$

where R is the resistance. In fact, a perfect resistor is always an LTI system for any setpoint (V_0, I_0) and for any frequency and amplitude. For actual resistors this will only hold for relatively small input and output signals:

$$\delta I(t) = \frac{1}{R} \delta V(t)$$

Taking the Fourier transform of this system equation gives:

$$\delta I(\omega) = \frac{1}{R} \delta V(\omega)$$

Hence the admittance $Y(\omega)$ is equal to $1/R$ and the impedance $Z(\omega)$ is equal to R for all frequencies ω . The phase of the admittance and the impedance is zero, since both system

functions are real. This is in line with the fact that the system equation of the resistor is an instantaneous, functional relation. The response of an ideal capacitor is described by the following differential equation:

$$I(t) = C \frac{dV}{dt}$$

where C is the capacitance. As the resistor, the capacitor is an LTI system for any setpoint, frequency and amplitude. Again for actual capacitors, this will only hold for relatively finite input and output signals:

$$\delta I(t) = C \frac{d(\delta V)}{dt}$$

Taking the Fourier transform of the system equation now gives:

$$\delta I(\omega) = i\omega C \delta V(\omega)$$

Hence the admittance of a capacitor reads $Y(\omega) = i\omega C$ and the impedance reads $Z(\omega) = 1/i\omega C$. It follows that the admittance and impedance of a capacitor depend on frequency, in contrast with a resistor.

Note that the system functions $Y(\omega)$ and $Z(\omega)$ of both the resistor and the capacitor obey property (3.4) of real systems. This also holds for RC-networks, composed from resistors and capacitors. This follows from the fact that for two network elements in series, the impedances can be added while for two parallel network elements, the admittances can be added. And if property (3.4) is obeyed for two network elements separately, it is also obeyed for the sum of the network elements.

Constant phase elements

It is interesting that property (3.4) implies that not simply every system function is possible. For instance, the impedance $Z(\omega) = iR$ or the admittance $Y(\omega) = \omega C$ do not obey property (3.4) and do therefore not refer to real systems. On the other hand, one can also give examples of system functions that do satisfy property (3.4), but for which one cannot so easily write a physical model or system equation. This holds for the constant phase element (CPE), for which the impedance reads:

$$Q_\alpha(\omega) = (i\omega\tau)^\alpha$$

Here the parameter α is a real number between -1 and $+1$. Only for the integer values $\alpha = -1$, 0 and $+1$, for which the CPE is equivalent to a capacitor, a resistor and an inductor respectively, the CPE refers to a known physical system. Nevertheless, the CPE also satisfies (3.4) for non-integer values of α , as follows from writing the impedance of a CPE as:

$$Q_\alpha(\omega) = (i\omega\tau)^\alpha = (\omega\tau)^\alpha i^\alpha = (\omega\tau)^\alpha \exp(i\alpha \frac{\pi}{2})$$

It follows that the modulus and phase of a CPE are equal to $(\omega\tau)^\alpha$ and $\alpha\pi/2$ respectively. We see that the phase of a CPE is independent of frequency, which is the reason that the CPE carries its name. The complex conjugate of the CPE reads:

$$Q_\alpha^*(\omega) = (\omega\tau)^\alpha \exp(-i\alpha\frac{\pi}{2}) = (-i\omega\tau)^\alpha = Q_\alpha(-\omega)$$

And therefore the CPE obeys property (3.4) and refers to a real LTI system. This fact could be put aside as a peculiarity if it were not that CPEs actually seem to occur in many actual systems and are in fact commonly used for explaining measured impedance spectra in spite of the absence of a supporting physical model.

Two systems where CPEs are observed should be mentioned here. These are the interface between noble metals and electrolytes and dielectric materials. Not surprisingly many efforts have been spent in order to explain the CPE behavior for these particular cases. For metal-electrolyte systems, the fractal structure of the electrode surface (cf. [Jonscher 1983]) and the influence of the electrode geometry are mentioned as possible explanations (cf. [Wagner 1951, Hine 1956, Levie, de 1963]). For dielectric materials, the CPE response was already observed in the middle of the 19th century by Kohlrausch, who noticed that the discharge of capacitors with time could be described by response functions of the form $\exp(-at^b)$ where a and b are constant parameters [Kohlrausch 1854]. This was overshadowed by the work of Debye [Debye 1912], but rediscovered by Williams and Watts who also pointed out the relation with CPEs [Williams 1971]. In between, the brothers Cole had already proposed the expression $Q_\alpha(\omega) = (i\omega\tau)^\alpha$ for explaining the behavior of several classes of dielectrics [Cole 1941].

However, none of these authors could give some kind of physical explanation of their observations. Jonscher suggested that the CPE behavior of dielectrics could be due to the mutual interactions between the dipoles in dielectrics [Jonscher 1975]. Since then, several authors worked on this idea (cf. [Dissado 1979]), but so far no complete and convincing theory has been given. Worth mentioning is perhaps the result of Cahn from the theory of crystallization, who showed that during certain types of crystallization processes, the movement of the crystallization front is described by $\exp(-at^b)$ where a and b are parameters depending on the dimensionality of the crystallization process [Cahn 1996]. It is perhaps no coincidence that this is exactly the response function of Kohlrausch and perhaps the orientation of dipoles in dielectrics may also be described by the progress of a sort of orientation front, where the dipoles are orienting the following neighbors. Otherwise, the following mathematical models lead to CPE behavior and perhaps one of them may eventually be connected to a physical model (cf. [Jonscher 1983]).

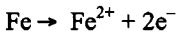
- Perhaps the response of a system is not described by a well-behaved function. If this is not the case, such as when the response essentially has a fractal structure, fractional derivatives are required to describe the system. An example of a fractional derivative is $d^{1/4}x/dt^{1/4}$. The Fourier transform of this term reads $(i\omega)^{1/4}X(\omega)$, leading to a CPE with non-integer $\alpha = 1/4$;
- Suppose the input and output signals not only depend on the time, but also on position. The system is then described by a partial differential equation such as for instance Fick's second law, which was discussed in chapter 2. This law applies to the electric current from an electrode into an electrolyte, in case the diffusion of ions through the electrolyte

is rate determining. It can be shown that the simultaneous presence of the first derivative to the time and the second derivative to the position in Fick's second law leads to the so-called Warburg impedance, which has the form of a CPE with $\alpha = -\frac{1}{2}$ (cf. [Jonscher 1983, Macdonald 1987]);

- Finally, CPE behavior can also be obtained by very large and well-balanced RC-networks. This is for instance done in the theory of explaining dielectric spectra by time constant distributions (cf. [Böttcher 1978, Jonscher 1983] and chapter 7).

Double layers

The origin of the electric impedance of a metal-electrolyte interface is the difficulty of charge carriers to cross the interface. In fact, for water-based electrolytes, there is even no charge carrier that can both be transported in the metal and the electrolyte. In the metal, the charge carriers are the electrons and in the electrolyte, the charge carriers are ions. For the propagation of an electric current through a metal-electrolyte interface, the electrons therefore have to be "modified" into ions. This is done by electrochemical reactions such as:



This dissolution reaction provides a positive current from the metal to the electrolyte. For this reason, this reaction is called an anodic reaction. The opposite reaction of the precipitation of iron ions is a cathodic reaction. Now when an alternating voltage is applied over the interface, iron will alternately dissolve and precipitate and an alternating current will flow, which is the response to the applied voltage.

The relation between the voltage and the current for a metal-electrolyte interface is a nonlinear differential equation, which in most cases will show a linear response to small signals. For low frequencies, the current response is dominated by the limited speed of the chemical reactions on the interface. We will first consider the low frequency case and later discuss the case for high frequencies.

For very low frequencies, all derivatives in the nonlinear system equation vanish and the current responds instantaneously to the voltage. This implies a functional relation between the current and the voltage:

$$I = f(V)$$

This steady state relation between the current and the voltage is called the "polarization curve" and is characteristic for the concerned metal-electrolyte system (cf. [Bockris 1972]). Now when a small variation δV is applied to this nonlinear system we have:

$$I_0 + \delta I = f(V_0 + \delta V) = f(V_0) + f'(V_0)\delta V + \frac{1}{2}f''(V_0)\delta V^2 + ..$$

where a Taylor expansion of $f(V)$ is carried out around the setpoint (V_0, I_0) . After the usual cancellation of the setpoint, we get the following in first order:

$$\delta I \cong f'(V_0)\delta V \cong \frac{1}{R_{ct}} \delta V$$

This is a linear system equation, provided that the first derivative in the setpoint $f'(V_0)$ does not vanish. The inverse of $f'(V_0)$ is then the impedance of the system and is called the “charge transfer resistance”. Hence we have:

$$Z(\omega) = R_{ct} = \frac{1}{f'(V_0)}$$

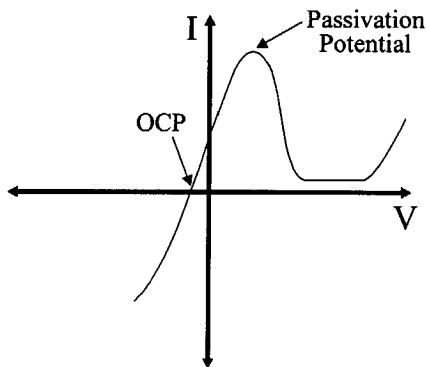


Figure 3.1. Polarization curve for a passivating metal-electrolyte system. At the open circuit potential (OCP), the net current through the circuit is zero. Beyond the passivation potential, the current starts to decrease with the voltage.

It follows that for low frequencies, the metal-electrolyte interface behaves as a resistor. For some metal-electrode systems, it is possible that $f'(V_0)$ is negative or even vanishes in certain setpoints. This holds for passivating systems, for which a typical polarization curve is sketched in figure 3.1. It follows that at the passivation potential, the system will at least be quadratic and shows a purely nonlinear response. However, EIS measurements are mostly done at the open circuit potential (OCP), where the net current through the interface is zero. There the current response will be linear for small signals.

For high frequencies, the polarization curve no longer gives a complete description of the system (as it is however sometimes suggested, cf [Darowicky 1995]) and the dynamics of the double layer must be considered. The double layer consists of two layers of opposite charge on either side of the metal-electrolyte interface, origination from the chemical potential difference of the charge carriers between the metal and the electrolyte. In chapter 5, a discussion on double layers can be found, as well as of course in standard works on electrochemistry (cf. [Bockris 1972]). The current response of the double layer is also nonlinear (cf. [Schmickler 1996]), but for small signals can be described by a capacitor C_{dl} , which is called the “double layer capacitor”. Since the charge transfer resistance R_{ct} is a

resistance between the “plates” of the double layer, R_{ct} is a resistance parallel to C_{dl} , as shown in figure 3.2.

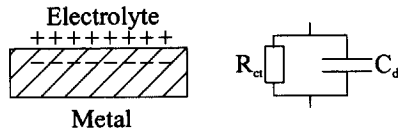


Figure 3.2. *Metal-electrolyte double layer and equivalent circuit for the linear response to small signals.*

It follows that at low frequencies, the major part of the current flows through R_{ct} and the response is indeed described by the polarization curve. For higher frequencies however, the charging and discharging of the double layer becomes dominant. We should add that the impedance of metal-electrolyte interface can be considerably more complicated, for instance when the adsorption of molecules on the surface has to be considered or when ion diffusion is rate determining (cf. [Bockris 1972]). In our case we will however see that the above description is sufficient.

3.3 Measuring the adhesion of coatings on metals by EIS

We will now discuss the application of electric impedance spectroscopy (EIS) for studying the delamination of polymer coatings from metals under the action of an electrolyte. For that purpose, we will first briefly discuss the delamination mechanism and then modeling the system by its electrical and electrochemical properties. The loss of adhesion then follows from the changes in some of the model parameters.

3.3.1 Delamination of polymer coatings from metals in electrolytes

Several mechanisms are mentioned for the loss of adhesion (or delamination) between polymer coatings and metals under the action of an electrolyte (cf. [Wit, de 1995]). The natural start is the transport of electrolyte through the coating. When the electrolyte has reached the metal surface, it can work on the metal-coating interface, for instance by dissolving the metal. Then the delaminated area grows and so on. (The diffusion of water may be accelerated by osmosis when salt products are present underneath the coating. In that case the water and salt products will form an electrolyte underneath the coating and the transport of ions through the coating is not necessary to form an electrolyte. Delamination will then take place much more rapidly. However, when the metal is properly cleaned prior to the application of the coating, this will not happen.)

The simultaneous electrolyte transport through the coating and the delamination by the action of the electrolyte on the metal-coating interface is a rather complex process. Moreover the transport of the electrolyte through the coating is often very slow, particularly

for the diffusion of the ions as we have seen in chapter 2. Hence one may have to wait a long time before delamination actually occurs. One would prefer to skip the diffusion of the electrolyte through the coating and to immediately measure the delamination process by the action of the electrolyte.

One way to do this is to create an artificial defect in the coating [Bonora 1988]. When the defect is not too small, it will almost instantaneously be filled by electrolyte and the diffusion process is skipped. Crucial for this method is of course that the artificial defects are well-defined. This leads to the following demands:

- The defects must have a well-defined and reproducible shape and size;
- The metal-coating system at the boundaries of the defect must not be damaged and be representative of the rest of the sample.

In fact, it is not so obvious to produce defects that are qualified for our purposes. For mechanically produced defects, such as by cutting with a knife or by puncturing with a needle, the coating is not really removed but deformed and the shape, size and depth of the defect are not easily controlled. Better results can be obtained by laser ablation using Excimer lasers [Srinivasan 1987, Reyna 1994]. This is however laborious and requires, besides an Excimer laser, some optical equipment such as at least a beam expander to improve the quality of laser beam [Glover 1995]. In the experimental section 3.4.1, our attempts of producing artificial defects are discussed.

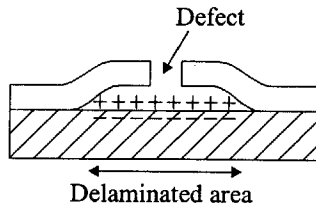


Figure 3.3. Progressed delamination started from an artificial defect in a coating. Drawn is the double layer at the delaminated area.

3.3.2 Characterization of coated metals in electrolytes

The technique of electric impedance measurements for the characterization of polymer coatings on metals dates in fact from the water uptake measurements of Hartshorn and Brasher around 1950, as mentioned in chapter 2. These were usually electric impedance measurements at a single or at a few frequencies around 1000Hz. In the 1970s, much better equipment for impedance measurements became available and impedance measurements of wide frequency spectra could be recorded much more easily and much more accurately. This opened perspectives for a more detailed characterization of polymer coatings on metals. Apart from the water uptake of intact coatings, which could already be studied from single frequency measurements, it became also possible to study the porosity of coatings and the contact area of defects where the metal is in contact with the electrolyte. For a review article

on the possibilities of EIS for the characterization of coated metals, the reader is referred to the paper of Kendig and Scully or the paper of Mansfeld [Kendig 1990, Mansfeld 1995].

Here we will focus on coated metals with an artificial defect. In figure 3.3, an artificial defect in a polymer coating on a metal is drawn. In this case, delamination has already increased the delaminated area considerably.

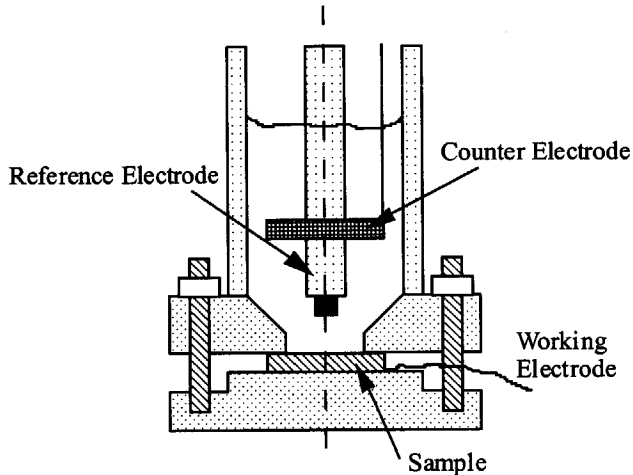


Figure 3.4. Setup for EIS measurements with three electrodes. The sample is clamped in the cell and a round area with a diameter of 2cm is immersed by electrolyte.

Now by an EIS measurement, the impedance between the metal underneath the coating and an electrode in the electrolyte is measured. This is normally done by a setup as sketched in figure 3.4. In this setup, three electrodes are used:

- The working electrode (WE), which is the metal underneath the coating;
- The counter electrode (CE) in the electrolyte which delivers the alternating current for the impedance measurement. This current runs through the CE and the WE;
- The reference electrode (RE). This is an electrode with a very well defined electric potential, such as a Saturated Calomel Electrode (SCE). The alternating voltage of the impedance measurement is measured between the RE and the WE.

In fact, also a system with two electrodes would be possible. Then the RE and the CE are one and the same and for instance the RE could be left out of the setup. The combined CE and RE is then both used for delivering the current and as a reference electrode. However the advantage of splitting up these electrode functions is basically that the RE then does not deliver a current and that the impedance contains no contribution of possible polarization effects of the RE. In addition, a very stable RE such as an SCE can be used so that the setpoint voltage and current can be set very accurately. The SCE is however not suited for following higher frequencies and therefore the SCE is usually coupled via a capacitor to a

platinum wire that is mounted besides the SCE. The platinum wire then follows the higher frequencies and a “dual reference electrode” is obtained. (For a detailed discussion this subject, the reader is referred to the article of Herrmann, Perrault and Pilla [Herrmann 1968].) The CE finally is a platinum mesh with a relative large surface.

With the setup of figure 3.4, the electric impedance between the RE and the WE is measured. Again, one can apply an alternating voltage and measure the resulting current or do the opposite. Supposing that the first is done, we have:

$$V(t) = V_0 + \delta V \cos(\omega t)$$

This is the applied voltage between the RE and the WE, where V_0 is the setpoint voltage and δV is the amplitude of the perturbation with angular frequency ω . The current between the RE and the WE is equal to the current from the CE to the WE (rather, this current is just passing the RE):

$$I(t) = I_0 + \delta I \cos(\omega t + \phi)$$

where I_0 is the setpoint current and ϕ the phase shift between the current and the voltage. The modulus of the electric impedance Z between the RE and the WE is now the ratio of δV and δI and the phase of Z is equal to the phase shift ϕ , or:

$$Z = |Z| \exp(i\phi_z) = \frac{\delta V}{\delta I} \exp(i\phi)$$

We will not discuss the clever electronics used for EIS measurements and assume that the above voltage and current can be applied and measured accurately. (For a discussion on the electronics of EIS, cf. [Macdonald 1987].) It is not always necessary to use a three electrode system. A three-electrode system is particularly necessary for nonlinear systems such as metal-electrolyte interfaces. For measurements on intact coatings however, such as in chapter 2, the setpoint (V_0, I_0) is irrelevant and a two electrode system suffices.

Now we will discuss the electric impedance of the system of figure 3.3, as mounted in a setup such as figure 3.4. The WE is then the metal underneath the coating and the RE is positioned in the electrolyte above the coating. For such systems, Menges and Schneider proposed the equivalent network of figure 3.5 to characterize the electric impedance [Menges 1973]. A few years later, more or less the same methods were developed independently by Beaunier, Epelboin, Lestrade and Takenouti [Beaunier 1976]. Nowadays, the equivalent circuit of figure 3.5 is widely used. As we will see, the components of the equivalent circuit can all be interpreted meaningfully. Going from the RE to the WE, the current passes respectively:

- The electrolyte, characterized by a resistor R_e ;
- The coating, characterized by a coating capacitor C_c and parallel to it, the resistance of the artificial defect R_d ;
- After the coating, the current already arrives at the WE. After the defect, the current first passes the metal-electrolyte double layer, which is characterized by a charge transfer resistance R_{ct} and a double layer capacitor C_{dl} .

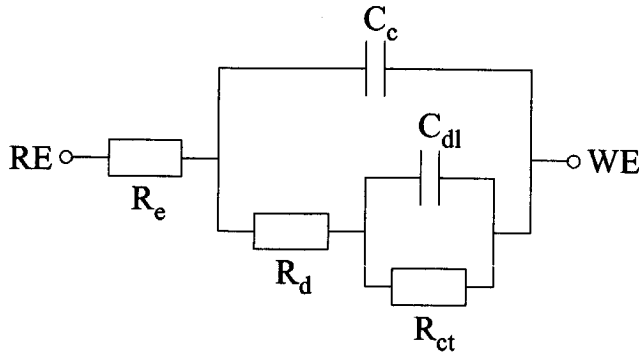


Figure 3.5. Equivalent circuit of the impedance between the RE and the WE for a metal with a porous coating or for a coated metal with a defect.

Actually, the equivalent circuit of figure 3.5 was originally designed for porous coatings with no artificial defect. Then R_d has a different meaning, being the resistance of the pores, but the equivalent circuit is the same.

The electric impedance spectrum of the network of figure 3.5 is plotted in figure 3.6. Only the modulus $|Z|$ is plotted. The curve is characterized by a number of breakpoint frequencies and within the frequency intervals, the impedance is dominated by different components of the network. The positions of the breakpoints depend on the magnitudes of the components. In case of a significant but small defect, one normally has $R_{ct} > R_d > R_e$ and $C_{dl} > C_c$ and the breakpoint frequencies read:

- $\omega_1 = 1 / R_{ct}C_{dl}$;
- $\omega_2 = 1 / R_dC_{dl}$;
- $\omega_3 = 1 / R_dC_c$;
- $\omega_4 = 1 / R_eC_c$;

Below the lowest frequency ω_1 , both capacitors are blocked and the current runs through the three resistors in series. The network then behaves as a resistor and $|Z| = R_e + R_d + R_{ct} = R_1$. Since R_{ct} is the largest of the three, it dominates the impedance in this frequency interval. Between ω_1 and ω_2 , the larger capacitor C_{dl} is shortcut and dominates the impedance. In the interval between ω_2 and ω_3 , the impedance of C_{dl} becomes negligible and the impedance is dominated by the sum $R_e + R_d = R_2$, or in approximation by only R_d , which is the larger resistor. Between ω_3 and ω_4 , the whole branch of the defect is shortcut by the coating capacitor C_c , which then dominates the impedance. This frequency region is therefore ideal for characterization of the coating only, such as for water uptake measurements. Finally above ω_4 , the impedance of C_c becomes negligible with respect to the electrolyte resistance $R_e = R_3$.

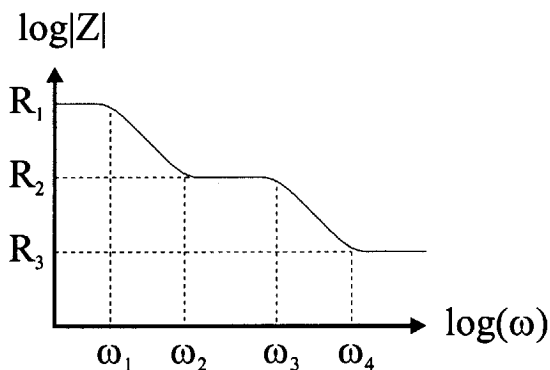


Figure 3.6. Typical impedance spectrum for a coated metal with a defect or for a porous coating under immersion of an electrolyte.

The double logarithmic plot of figure 3.6 is called a “Bode plot” and is particularly useful for representing impedance measurements, because one can clearly recognize when the coating behaves as a resistor or as a capacitor. For a resistor, the impedance is independent of frequency, which shows up as a horizontal line in the plot. For a capacitor on the other hand, $|Z| = 1/\omega C$, which leads in a double logarithmic plot to a straight line with slope -1 or in other words, descending under 45° .

Besides plotting $|Z|$ as a function of frequency, one can also consider the phase ϕ_Z of the impedance. The information from the phase is however not independent of the information from the modulus and one can even show that the whole phase plot $\phi_Z(\omega)$ can be derived from the modulus $|Z(\omega)|$ by for instance the Kramers-Kronig transform (cf. [Böttcher 1978]). However, measuring the phase plot is still useful, either for validation of the modulus data or as additional data for fitting the values of the components of the equivalent network.

It should be mentioned that also considerably more complicated equivalent networks than figure 3.5 are used (cf. Hirayama 1991). However, one should not try to extract too much information from the data. For instance, the final plateau for $\omega > \omega_4$ will not be visible when the electrolyte resistance is very small, and then this parameter should then be omitted. The addition of extra, exotic network elements is tempting because it always leads to a better fitting of the measurements. The result may then however be meaningless.

3.3.3 Characterizing the delamination of coatings

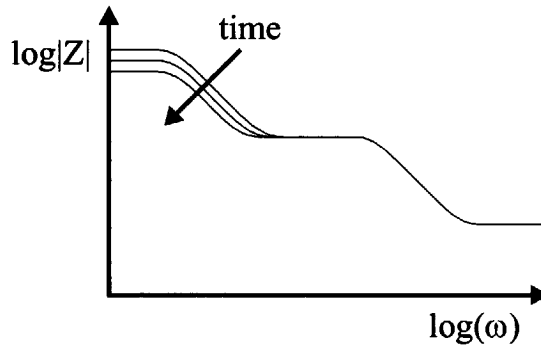
Now we come to the main idea of following the delamination of the coating by EIS and discuss the influence on the elements of the equivalent circuit of figure 3.5:

- The electrolyte resistance is approximately equal to $R_e = d_e / \gamma_e A_c$. Here, d_e is the distance between the RE and the coating (or the thickness of the electrolyte layer), γ_e is the conductivity of the electrolyte and A_c is the immersed area of the coating. It follows that R_e does not change during the delamination process;

- The coating capacitance reads $C_c = \epsilon_c A_c / d_c$, where d_c is the thickness of the coating and ϵ_c is the permittivity. C_c is not affected by the delamination process when the area of the coating A_c is much larger than the delaminated area A_{dl} ;
- The defect resistance can be estimated by $R_d = d_c / \gamma_e A_d$, where the thickness of the coating d_c is equal to the height of the defect and A_d is the area of the defect. This quantity will also not be affected by the delamination process;
- The charge transfer resistor R_{ct} is inversely proportional to the area of the double layer A_{dl} , which is equal to the delaminated area;
- The double layer capacitor is proportional to A_{dl} .

It follows that the delamination process only affects R_{ct} and C_{dl} . Hence by following R_{ct} or C_{dl} , the progress of the delamination process can be studied. (In addition, R_{ct} is inversely proportional to the corrosion current by the Stern-Geary relation [Stern 1957] and is therefore a measure for the rate of metal uptake of the electrolyte. This is interesting for the prediction of the metal uptake of the electrolyte, or of canned consumables [Bird 1971, Kleniewski 1975, Grandle 1994, Bovard 1995, Dolfing 1995, Seruga 1996, Galic 1998, Delucchi 1999]).

In terms of breakpoint frequencies, only $\omega_2 = 1 / R_d C_{dl}$ is affected and is inversely proportional to the delaminated area A_{dl} . The other breakpoint frequencies ω_3 and ω_4 are independent of A_{dl} and so is $\omega_1 = 1 / R_{ct} C_{dl}$, since R_{ct} is inversely proportional to A_{dl} and C_{dl} is proportional to A_{dl} . In figure 3.7, the changes of the Bode plot with time are shown during a delamination process at an artificial defect.



Figures 3.7. Bode plot of a coated metal with an artificial defect where delamination occurs from the boundaries of the defect.

This completes the description of the studying the delamination of coatings under the action of an electrolyte by EIS on artificial defects. The expected shape of the spectrum is described by figure 3.6 and the delamination process is parametrized by several elements of the equivalent circuit of figure 3.5. We will now finish the discussion by some remarks on the validity of the equivalent circuit:

- It is assumed that the electric current flows vertically from the CE to the WE and vertically through the defect. In reality however, the current pattern will be more complex, especially around the significant region of the artificial defect. The consequence is for instance that the top of the coating will not be an equipotential surface. However in the equivalent circuit, the top of the coating is only represented by one point and hence by only one potential. The current pattern will be different for different frequencies and the influence on the impedance spectrum is hard to estimate. Computer calculations have been done in the past for that purpose (cf. [Wagner 1951, Hine 1956, Levie, de 1963]) and may be helpful for future research, especially since the recent development in software applications such as Matlab®;
- The coating is modeled by a simple capacitor. The dielectric losses of the coating material, leading to an imaginary part of the coating permittivity (and to a real part of the coating impedance) are thus ignored;
- The double layer model of a charge transfer resistance R_{ct} parallel to a double layer capacitor C_{dl} can be too simple to describe the metal-electrolyte interface.

These issues may only have a small influence. For instance when the conductivity of the electrolyte is very high, the whole electrolyte can be regarded as an electrode at one and the same potential and the influence of the current pattern is very small. And for delamination measurements as described above, there is no need to describe the impedance of the coating very accurately. Whether the double layer model with R_{ct} and C_{dl} is valid depends of course strongly on the metal-electrolyte system.

3.4 The delamination of PET coated ECCS in citric acid

We will now discuss a number of measurements of the delamination rate of PET coatings on ECCS in citric acid. First we will discuss the creation of the defects, then the metal-electrolyte system and finally the delamination measurements by EIS.

3.4.1 Creating artificial defects

We created the artificial defects with a needle. The needle was mounted to a motion-controller and could be moved with a precision of $1\mu\text{m}$. We punctured the needle through the coating until the needle made electric contact with the metal underneath. Another way of making artificial defects by a needle is to load it with a well-defined weight [Bonora 1988]. An advantage of our method is that one is sure that the coating is locally removed and that the electrolyte will make contact with the metal.

Puncturing defects is a rapid and cheap way of producing defects. However as shown in figure 3.8, the coating is deformed around the defect. The refractive index of PET is changed at the deformation, which shows up as a ring around the defect. It follows that puncturing by a needle is not an ideal way of creating artificial defects, since we will measure the delamination of the deformed coating on the boundaries of the defect.



Figure 3.8. Artificial defect of about $70\mu\text{m}$ in PET coated ECCS created by puncturing with a needle. The photograph was made by an optical microscope.

A probably better way of creating artificial defects is by laser etching with Excimer lasers. This has been done previously for corrosion studies on polymer coated metals (cf. [Geenen 1990, Koot 1991]). By choosing the right wavelength of the laser light and the right range of the intensity of the laser beam, the coating is removed without melting, which process is called “laser ablation” (cf. [Garrison 1984, Srinivasan 1986, Srinivasan 1987, Znotins 1987, Reyna 1994, Glover 1995]). In figure 3.9, the influence of the energy density is illustrated. It is shown that the energy density must be high enough to evaporate the polymer fragments, but low enough to prevent melting of the coating at the edges of the defect. The energy range for ablation is roughly a factor 20 wide (cf. [Srinivasan 1987, Reyna 1994, Glover 1995]).

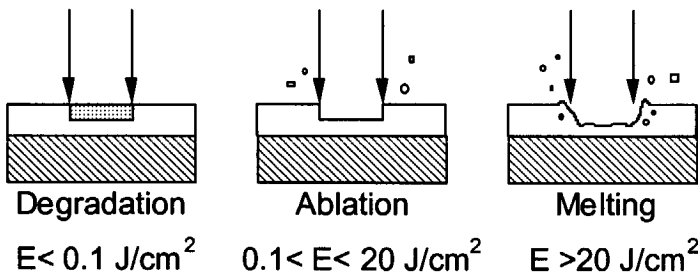
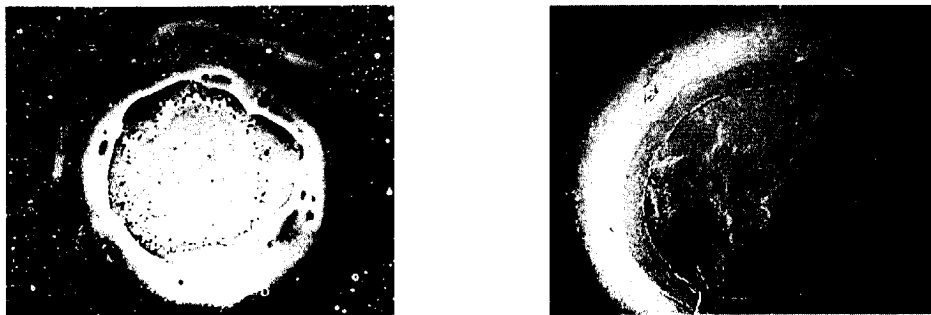


Figure 3.9. Creation of artificial defects by an Excimer laser. The energy ranges are taken from Glover, Illy and Piper [Glover 1995].

In figures 3.10 and 3.11, the influence of the energy density on the resulting defect is illustrated. We see that at high energies, the coating has melted around the defect and the metal-coating system around the defect is not well defined. In the ablation region, the walls of the defect have not been molten and very smooth.



Figures 3.10 and 3.11. Laser defects created by a too high energy density and by ablation. The defect of figure 3.10 was created by using the third harmonic of a YAG-laser (355nm) and the defect of figure 3.11 by a XeCl Excimer laser (308nm). The photographs are made by a scanning electron microscope (SEM).

Besides the intensity of the laser beam, the wavelength of the laser light is also of importance. For the creation of a well-defined defect, the absorption of the laser light should be maximal to prevent heating of the metal underneath the coating before the defect is ready. The metal then would transport the heat to areas around the defect, by which the metal-coating system could be modified. In order to choose the right laser, we therefore measured the transmission of our PET films as a function of the wavelength. The results are shown in figure 3.12. We see that below 310nm the transmission of our PET films of 25 μm sharply drops to about 0.05%. This value is more or less maintained for wavelengths between 200 and 300nm.

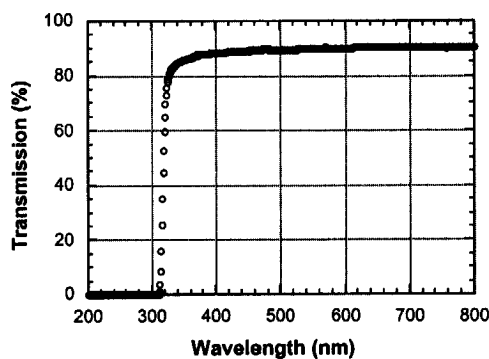


Figure 3.12. Transmission spectrum of a PET film of 25 μm . We see that between 310 and 320nm, the transmission rapidly drops.

It follows that of the 4 types of Excimer lasers that are commercially available, the ArF (193nm), KrF (248nm) and the XeCl (308nm) are suited for the creation of artificial defects in PET, while the XeF (351nm) is not. Instead of using an Excimer laser, one can also think of using a YAG laser (1065nm) and taking the fourth harmonic (266nm). The wavelength of the third harmonic of a YAG laser (355nm) is too high.

The size of laser defects is controlled by a lens system. It is best to put the coating in focus where the laser beam has a cylindrical cross section, leading to defects with vertical walls. The focal length should be larger than the thickness of the coating. Additional lenses may be necessary to improve the quality of the laser beam, for which purpose a beam expander and a diaphragm can be used (cf. [Glover 1995]). Excimer lasers are pulsed lasers that produce up to about 10 pulses per second with duration of about 50ns. In the ablation range, the coating material is removed with about 1 μ m per pulse. Ablation produces sound, and it can be heard when the coating is removed by a significant decrease of the sound level. The latter can also be observed by the increased reflection of the laser light when the laser has reached the metal.

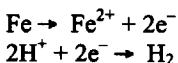
Laser defects seem to show better perspectives than mechanically produced defects. At the time when the experiments were done however, we had no good laser defects to our disposal and this is the reason why we worked with punctured defects. This means that they are not very reliable for measuring the actual delamination rate and the measurements below are therefore basically a test for the method.

3.4.2 Iron in citric acid

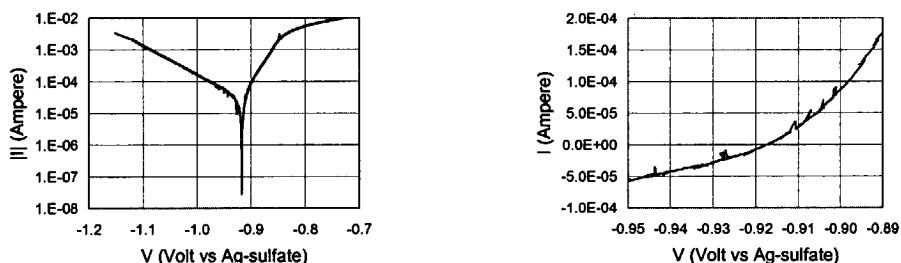
The electrochemical system in our artificial defects will be iron in citric acid, which is the chosen model electrolyte for consumables. The metal underneath the PET coating is actually ECCS, but this is iron with a chromium-chromium oxide layer of only a few nanometers, which main function is to provide a better adhesion of the coating. This electrochemically applied chromium-chromium oxide layer does show some corrosion protection and the initial behavior of ECCS in citric acid differs from iron in citric acid (cf. [Azzeri 1975]). However, since the chromium-chromium oxide layer is only very thin, it is not likely that it will provide significant corrosion protection during shelf life. Therefore, we ignored the chromium-chromium oxide layer for the moment. One liter of our model electrolyte contained:

- 38.43g of citric acid monohydrate;
- 4g of sodium hydroxide (NaOH);
- 0.5g of potassium sorbate.

This results in a buffered solution with a pH value of 3.0 at room temperature. The potassium sorbate is added to prevent the growth of microorganisms. For the simulation of delamination in food and beverage cans, oxygen should be removed from the electrolyte. The corrosion process is then relatively simple and described by:



These reactions are the dissolution of iron and the evolution of hydrogen gas. At the given pH value, the Pourbaix diagram of iron in citric acid (cf. [Kubal 1995]) indicates that the iron surface will be clean, without a passive layer. This is also supported by other workers (cf. [Uhlig 1963, Sekine 1990]) and by a measurement of the polarization curve of iron in buffered citric acid, shown in figures 3.13 and 3.14.



Figures 3.13 and 3.14. Polarization curve of iron in buffered citric acid of pH = 3, Tafel plot and linear plot around $I = 0$. The exposed area is 3.14cm^2 .

The Tafel plot of figure 2.13 shows clear cathodic and anodic Tafel-slopes. Analysis of the Tafel plot leads to the following results:

- A cathodic Tafel-slope of $b_c = 110\text{mV/decade} = 48\text{mV}$ per factor e ;
- An anodic Tafel-slope of $b_a = 39\text{mV/decade} = 17\text{mV}$ per factor e ;
- A corrosion current of $I_{\text{corr}} = 3 \times 10^{-5}\text{ A}$.

Here the corrosion current is determined from the intersection of the Tafel asymptotes. This quantity can also be determined from the Stern-Geary relation [Stern 1957]:

$$I_{\text{corr}} \left[\frac{1}{b_a} + \frac{1}{b_c} \right] = \frac{1}{R_{\text{pol}}}$$

Here, R_{pol} is the polarization resistance, which is defined as the slope dV/dI at $I = 0$. In fact, R_{pol} is equal to the charge transfer resistance at $I = 0$. When the Stern-Geary relation is written in this form, the Tafel slopes must be specified per factor e , which leads to a factor $\ln(10) = 2.3$. From figure 3.14, it follows that R_{pol} is about 300Ω , leading to $I_{\text{corr}} = 4 \times 10^{-5}\text{ A}$, in reasonable agreement with the above value.

The curves of figures 3.13 and 3.14 also contain the so-called ‘‘ohmic loss’’ of the voltage due to the electrolyte resistance between the RE and the double layer. However the conductivity of our buffered citric acid, containing large amounts of sodium, hydrogen and citric ions, is relatively high. A measurement gave the value of about $\gamma = 5 \times 10^{-3}\Omega^{-1}\text{cm}^{-1}$, which for an area of 3.14cm^2 and a distance of $1/2\text{cm}$ between the RE and the sample leads to $R_e = 30\Omega$. This estimate is both inaccurate (due to the simplification of the geometry) and

relatively small compared to R_{pol} (which is about 300Ω). Therefore we did not bother about correcting the polarization curve for the ohmic loss.

The measured Tafel-slopes are in good agreement with theoretical values for the anodic and cathodic Tafel-slopes for iron in deaerated acids, that are $b_c = 120\text{mV/decade}$ and $b_a = 60\text{mV/decade}$ for symmetrical potential pits with transfer coefficients $\alpha_c = \alpha_a = \frac{1}{2}$ (cf. [Uhlig 1963]). With enough confidence in the measurements, one can also fit the following transfer coefficients to the measured Tafel-slopes:

- $1-\alpha_c = (120/110) \times 0.5 = 0.55$, or $\alpha_c = 0.45$;
- $\alpha_a = (60/39) \times 0.5 = 0.64$.

The above values for the corrosion current and the polarization resistance depend on the size of the exposed area, which was 3.14cm^2 . Normalizing the values to this leads to $I_{\text{corr}} = 1 \times 10^{-5} \text{A/cm}^2$ and $R_{\text{pol}} = 940\Omega\text{cm}^2$, which values are characteristic for the interface between iron and buffered citric acid of $\text{pH} = 3$.

Finally figure 3.14 also gives an indication of the linearity of the system, which is useful information for choosing the voltage amplitude for the impedance measurements. We see that around $I = 0$, the polarization curve is approximately a straight line within a voltage range of about 10mV , hence for amplitudes ΔV below 10mV , one may expect the system to respond linearly. However the polarization curve does not contain information of the linearity for higher frequencies (whether modeling by the double layer capacitor C_{dl} is valid or not). Therefore the curvature of the polarization curve is no more than an indication for the linearity of the system.

3.4.3 Delamination measurements on PET coated ECCS

With punctured defects with a diameter of about $70\mu\text{m}$ we did measurements on the delamination of PET coated ECCS in buffered citric acid of $\text{pH} = 3$. As described above, the defects were not really qualified for the method, since the PET coating at the edges of the defects is damaged. Therefore these measurements are rather a test for the method than actual measurements of the adhesion of PET on ECCS and we did not bother about removing the oxygen from the electrolyte, as it should actually be done for imitating the environment in food and beverage cans.

Prior to electrochemical measurements, one often has to wait some time before the metal-electrolyte system is stabilized (since measuring on a rapidly changing, not time invariant system is undesirable.) For the rather simple case of iron in citric acid, stabilization takes place quite rapidly after immersion. But for small artificial defects in PET coatings, it can take some time before the electrolyte has entered the defect, due to the hydrophobic behavior of PET. As a measure for the stabilization of the system, the open circuit potential (OCP) is often considered. In figure 3.15, this quantity is plotted as a function of time for one of our defects of $70\mu\text{m}$. This figure shows that for defects in PET of this size, wetting takes place relatively rapidly.

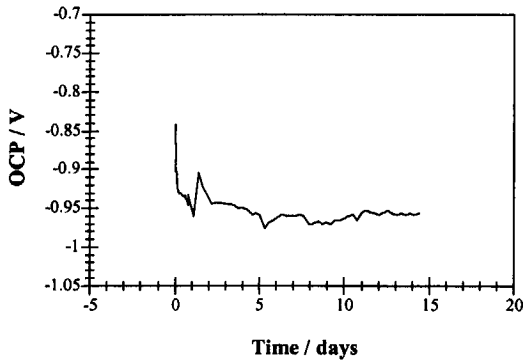


Figure 3.15. Stabilization of the OCP of a punctured artificial defect with a diameter of $70\mu\text{m}$ in PET coated ECCS in buffered citric acid of $\text{pH} = 3$.

For smaller defects in PET however, typically below about $50\mu\text{m}$, wetting of the defects can be a practical problem. In that case, the addition of surface-active agents may help to overcome the hydrophobic behavior of PET, such as for instance by the addition of a fraction of 1/500 of dishwashing product to the electrolyte. (We also found that the addition of surface-active agents is useful for improving the resolution of the well-known technique of detecting defects in coatings by immersion in a CuSO_4 -solution. In fact, we found rapid precipitation of copper in defects in PET coated ECCS as small as $5\mu\text{m}$ [Jong, de 1997].)

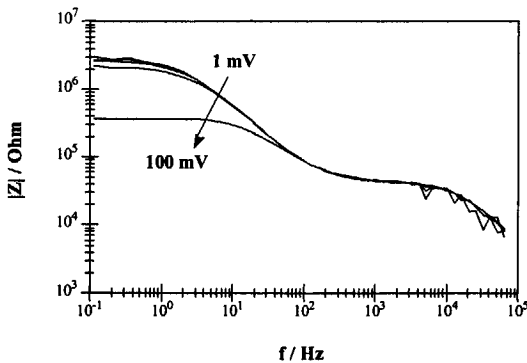


Figure 3.16. The influence of the amplitude on EIS measurements of artificial defects in PET039. The amplitudes are $\Delta V = 1\text{mV}$, 3mV , 10mV , 30mV and 100mV .

For the EIS measurements, the samples were mounted in the clamp cell of figure 3.4. The reference electrode was a dual AgSO_4 electrode, with its tip at $\frac{1}{2}\text{cm}$ above the sample. A Solartron 1250 Frequency Response Analyzer was used for the impedance measurements. Bandwidth C was chosen. All measurements were carried out at the open circuit potential (OCP). After some tests, a frequency spectrum between 100mHz and 65kHz was chosen together with an amplitude $\Delta V = 10\text{mV}$. Figure 3.16 motivates these choices. In this figure, the impedance spectra of an artificial defect are shown for amplitudes of 1mV, 3mV, 10mV, 30mV and 100mV. We see that the spectrum of 1mV contains random errors due to noise while the spectra for 30mV and 100mV are systematically different due to the non-linearity.

From the figure, we estimated that for 10mV and below the influence of the amplitude on the impedance spectrum was negligible and the system responded linearly. The non-linearity of the low frequency part indicates that at these frequencies, the impedance is indeed dominated by the metal-electrolyte interface. It also follows from figure 3.16 that the lower bound of the frequency spectrum of 100mHz is low enough to see the essential breakpoints in the impedance spectrum. These are about:

- $f_1 = 3\text{Hz} = 1/2\pi R_{ct}C_{dl}$;
- $f_2 = 200\text{Hz} = 1/2\pi R_dC_{dl}$;
- $f_3 = 10\text{kHz} = 1/2\pi R_dC_c$.

These are the breakpoint frequencies corresponding to equivalent circuit of figure 3.5 and the model spectrum of figure 3.6, except for the fact that the fourth breakpoint f_4 is missing. Above f_4 , the electrolyte resistance R_e dominates the impedance, but apparently R_e is too small to appear in the spectrum below 65kHz. This agrees with the earlier estimate of 30 Ω for the electrolyte resistance. The spectra can be matched to the following values for the parameters of the equivalent circuit:

- $R_{ct} = 3\text{M}\Omega$;
- $C_{dl} = 20\text{nF}$;
- $R_d = 40\text{k}\Omega$;
- $C_c = 300\text{pF}$.

The value of the coating capacitance C_c is in agreement with the measurements of chapter 2 for intact coatings. Also R_{ct} and R_d can be estimated. From the earlier measurements on iron in buffered citric acid we found the value of $R_{pol} = 940\Omega\text{cm}^2$ for the polarization resistance. Since we measured at OCP, we have $R_{ct} = R_{pol}$ and assuming that the area of the double layer is equal to the area of the defect having a diameter of $70\mu\text{m}$, we arrive at $R_{ct} = 20\text{M}\Omega$. The measured value is smaller, which may be due to the oxygen content of the electrolyte. An estimate for R_d follows from considering the defect as a cylinder with a diameter of $70\mu\text{m}$ and with a height of $25\mu\text{m}$, which is equal to the thickness of the coating. Supposing that this cylinder is filled by buffered citric acid with a conductivity of $\gamma = 5 \times 10^{-3}\Omega^{-1}\text{cm}^{-1}$, we arrive at a value of $R_d = 13\text{k}\Omega$. Comparison of these estimates to the above values shows that the estimates are quite rough, but have about the right order of magnitude.

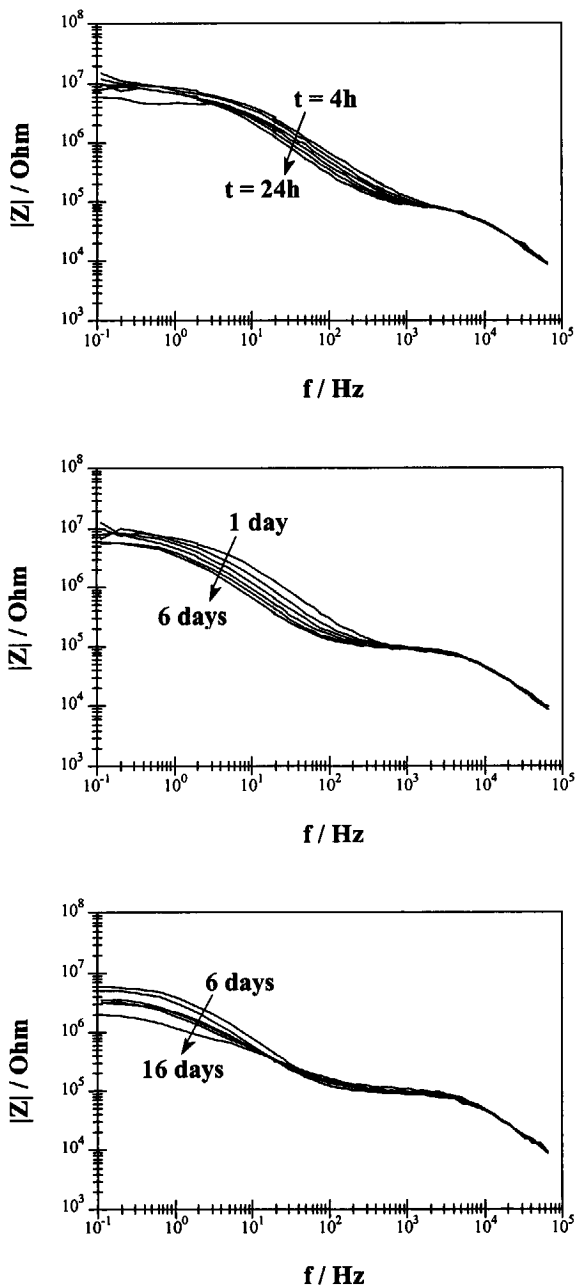


Figure 3.17. Impedance spectra as a function of time for a defect in PET039.

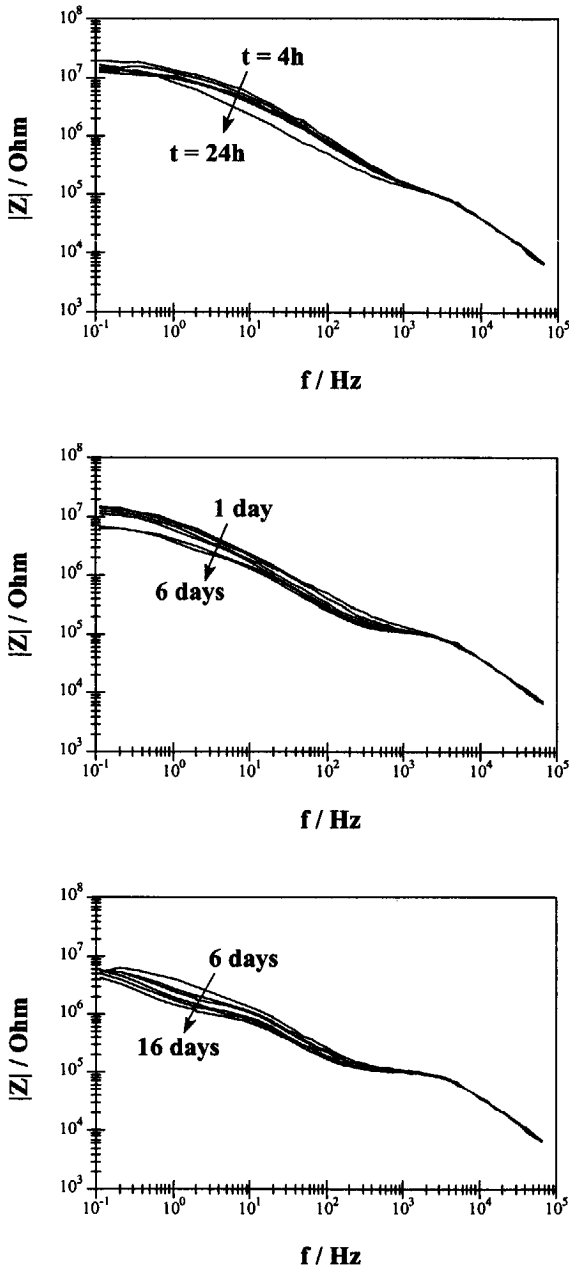


Figure 3.18. Impedance spectra as a function of time for a defect in PET052.

We now arrive at the original goal of the project, which is to follow the model parameters R_{ct} , C_{dl} , R_d and C_c as a function of time in order to follow the long-term delamination process of the coating. As discussed in section 3.3.3, it was expected that R_{ct} would decrease with the delaminated area A_{dl} while C_{dl} would increase by the same factor. On the other hand it was expected that R_d and C_c would remain more or less constant since the size of the defect does not change and the coating is rapidly saturated with water (in about half an hour as discussed in chapter 2). In terms of breakpoint frequencies, this implies that $f_1 = 1/2\pi R_{ct}C_{dl}$ and $f_3 = 1/2\pi R_d C_c$ would remain constant, while $f_2 = 1/2\pi R_d C_{dl}$ would decrease with $1/A_{dl}$.

Indeed the impedance measurements of figures 3.17 and 3.18 seem to support these ideas and with that the validity of the equivalent circuit and the method of following the delamination of coatings from metals under the action of an electrolyte. Shown are the changes of the impedance spectrum from the first day until the sixteenth day after immersion. We measured on to types of PET, labeled PET039 and PET052 by Corus RD&T.

Figures 3.17 and 3.18 show a fair agreement with figure 3.7. The direction in which the figures shift with the time agrees, as well as the fact that the higher frequency part ruled by R_d and C_c virtually does not change with time. However some significant deviations may also be observed. Particularly figure 3.18 for the defect in PET052 differs and even seems to show an additional breakpoint at low frequencies. Several reasons can be mentioned why the measurements might disagree with figure 3.7:

- The punctured defects were not qualified for our purposes;
- The impedance of the metal-electrolyte interface is more complicated than expected and cannot be described by only a charge transfer resistance R_{ct} and a double layer capacitor C_{dl} . This could be due to the oxygen not being removed from the electrolyte;
- The sulfate dual electrode, supplied by Corus RD&T and used as a reference electrode (RE) is perhaps too slow for impedance measurements. Then in a finite frequency range, the RE cannot follow the alternating voltage, resulting in a too low impedance value. This is subject of future research;
- The dielectric losses of PET are discarded, but this is not so important here;
- The influence of the geometry is ignored. Since the conductivity of the electrolyte is rather high, it is maybe not so important that the electric field lines are not the same for all frequencies. However the actual effect is difficult to estimate.

All these issues ask for future research and improvement. The method of working with artificial defects however shows rather good perspectives, especially when working with relatively simple systems as iron in deaerated citric acid.

3.5 Conclusions and suggestions for future research

We have investigated the abilities of using EIS measurements for determining the loss of adhesion of organic coatings from metals by the action of electrolyte. This is not really a new method, but normally one works with coatings as delivered by the manufacturer. Then one not only measures the delamination of the coating, but also the transport of the electrolyte through the coating. And in case the latter process is very slow, such as for the diffusion through PET as discussed in chapter 2, one will virtually never come to studying the actual

delamination process. In order to skip the diffusion process, we created an artificial path for the electrolyte to the metal by means of an artificial defect in the coating. Then delamination starts right after immersion by the electrolyte.

However, it appeared to be rather difficult to create small and well-defined defects of which the edges of were intact. Defects produced mechanically do not qualify and also the creation of laser defects is not so obvious, since an UV-laser beam of high quality is needed, requiring specialized optical equipment. At the time of the experiments, no sufficiently good defects were available yet and we worked with punctured defects. Therefore the above experiments must rather be seen as a test for the method than as actual measurements of the loss of adhesion of PET coatings from ECCS in citric acid. Recently however, much better defined laser defects were produced at Acordis R&D and it is hoped that with these defects, more significant experiments can be done. The following is a list of additional suggestions:

- Work with laser defects;
- Optimize the size of the defects. Smaller defects lead to more rapid changes of the observed quantities, but also to larger impedance values that are more difficult to measure with high accuracy;
- When wetting of the defects is a problem, the addition of surface-active agents to the electrolyte can be considered;
- Remove the oxygen from the electrolyte. Thus, the environment inside a can is better imitated and in addition, a simpler electrochemical system is obtained;
- Study the influence of the geometry by computer calculations to vary the equivalent circuit. For instance the PDE-toolbox of Matlab V, by which the Laplace equation can be solved for complex variables, can be useful for that purpose.

When the method is well established, one can think of studying the influence of heat treatments and deformation on the adhesion of the PET coatings, as well as the influence of different additives to the electrolyte, the influence of the pressure and so on. Besides one can of course think of applying the method to other coated metals. However, the first priority is to create well-defined defects that meet all requirements for the method.

Finally one limitation of the method should be mentioned, which is the fact that only delamination under the immersion of liquid electrolyte can be observed. However, under a humid, salty environment, we also observed corrosion delamination in the form of filiform corrosion. (In fact, filiforms of up to 4mm were observed at some PET coated ECCS samples with artificial cutting edges when exposed to a humid KCl-environment.)

Chapter 4

Deformation of PET coated ECCS during a DWI-process

4.1 Introduction

In this chapter we present a number of measurements on the deformation of PET coatings on ECCS during a deep drawing and wall ironing process (DWI-process) for can manufacturing. We measured the local deformation of the can wall as a function of the height by creating round defects in the plane base material, which was coil coated PET on ECCS. The measurements were motivated by a discussion on the degree of deformation in the can wall and in particular whether or not the deformation of the can wall depends on the height. In short, the following factors were studied:

- The deformation of the can wall as a function of the height, hence the deformation of the round defects;
- The influence of the rolling direction on the final deformation;
- The PET type. Two types of PET were studied, namely PET039 and PET052;
- The intermediate deformation during the DWI-process.

The deformation measurements are compared to the predictions of a simple geometric model, which is explained in the following section.

4.2 Geometrical model for a DWI-process

We first determine the final position of a volume element of the plane sheet material in the can wall. It is assumed that:

- The can wall remains intact;
- The deformation is plastic, hence the density of the can material is constant;
- The thickness of the can wall is independent of the height.

The latter assumption may seem unreasonable since the lower part of the can wall is exposed to more stress than the higher part, but is verified by measurements on the thickness of the can wall, as shown in the following section.

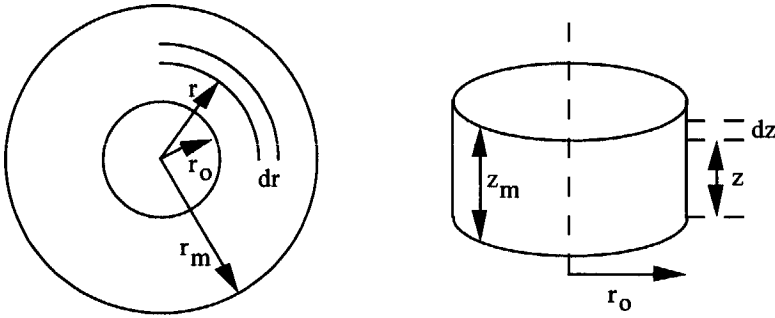


Figure 4.1. Parameters of the DWI-process. On the left, a plane sheet is shown and on the right a cup, created by deforming the plane sheet.

In figure 4.1, the parameters of the DWI-process are shown. Two additional parameters are the thicknesses a and b of the base material and the can wall. The distance between the punch and the mould determines the latter thickness. The final diameter of the can is r_0 .

During the process, a ring with width dr at a distance r from the center of the base material is deformed to a ring with width dz at height z in the can wall. The bottom of the can, for which $r \leq r_0$, is not deformed. (In practice, the bottom is actually deformed, since a profile is imprinted in the bottom to improve the strength of the can.) Since the density of the ring is constant, it follows that:

$$a \times 2\pi r dr = b \times 2\pi r_0 dz$$

From this equation, the width dz of a ring in the can wall follows:

$$dz = \frac{a}{b} \frac{r}{r_0} dr \quad (4.1)$$

The height z of the ring is obtained by integration of dz between r_0 and r :

$$z = \int_{r_0}^r dz = \frac{a}{b} \frac{1}{2r_0} [r^2 - r_0^2] \quad (4.2)$$

In particular we have the following for the total height of the can z_m :

$$z_m = \int_{r_o}^{r_m} dz = \frac{a}{b} \frac{1}{2r_o} [r_m^2 - r_o^2] \quad (4.3)$$

For very low cups, with $(r_m - r_o) \ll r_o$, the latter can be approximated by:

$$z_m \cong \frac{a}{b} (r_m - r_o)$$

For instance when $a = b$, we see that the final height of the can wall is equal to $(r_m - r_o)$ for very low cups, as one might intuitively guess from figure 4.1. The outer part of the disk with $r_m > r_o$ then simply "flips over" into the can wall.

We can also determine the deformation of a volume-element in the can wall. The deformation of the volume-element is characterized by:

- The deformation in the z-direction parallel to the axis of the cup;
- The deformation in the azimuthal direction around the cup;
- The reduction of the thickness of the can wall.

The deformation in the z-direction, or axial direction, follows from the relative change of the thickness of a ring dr in the plane sheet and the height dz of the corresponding ring in the can wall. The ratio of dz and dr follows from equation (4.1):

$$\frac{dz}{dr} = \frac{a}{b} \frac{r}{r_o} \equiv \varepsilon_z \quad (4.4)$$

When $\varepsilon_z = 1$, there is no axial deformation. However since $b \leq a$ and since $r > r_o$, we will always have $\varepsilon_z > 1$ in the can wall, hence the can material in the wall is always stretched along the axis of the can. In addition it follows that higher in the can, where r is large, the axial deformation is larger.

The deformation in the azimuthal direction follows from the reduction of the perimeter of a ring dr in the plane sheet when it is deformed into a ring dz in the can wall. In the plane sheet, the perimeter increases with r by $2\pi r$, while in the can wall, the perimeter is constant and equal to $2\pi r_o$. Hence the azimuthal deformation reads:

$$\frac{2\pi r_o}{2\pi r} = \frac{r_o}{r} \equiv \varepsilon_\theta \quad (4.5)$$

Since $r \geq r_o$ it follows that $\varepsilon_\theta \leq 1$, hence that the material in the can wall is always compressed and more compressed at greater heights.

Finally, the third deformation of a volume element in the can wall is the reduction of its thickness:

$$\frac{b}{a} \equiv \varepsilon_w \quad (4.6)$$

Note that the product of the three deformations of equations (4.4), (4.5) and (4.6) is always equal to 1, conform to the fact that the deformation is plastic.

4.3 Some measurements on cans during a DWI-process

At the Packaging Application Center (PAC) of Corus RD&T, a number of cans were made of PET coated ECCS by a DWI-process of 2 deep drawing steps and wall ironing with 3 rings (cf. [Aa 1999]). The diameter of the initial round sheet was 14.9cm. After the first deep drawing step a cup with a diameter of 10cm was produced, and after the second deep drawing step the final diameter of 7.5cm was reached. Wall ironing then finally increased the height of the can. Also during wall ironing, a profile is created in the bottom of the can. Intermediate cups after 1 or 2 deep drawing steps were produced, as well as after 1 ring of wall ironing. Together with the final result, this made four different formation steps to study.

Prior to the deformation process, round defects with a diameter of $70\mu\text{m}$ were created in the base material. The defects were the same punctured defects as in chapter 3. The defects were created at distances $r = 4.65, 5.20, 5.75, 6.30$ en 6.85cm from the center of the initially round sheet, as shown in figure 4.2. We created series of defects parallel to the rolling direction (denoted below by the symbol \parallel), perpendicular to the rolling direction (denoted by the symbol \perp) and under 45° with the rolling direction (denoted by the symbol \diagup). We created defects on both sides of the plane sheet in order to both study the deformation inside and outside the can wall. Finally, as mentioned in the introduction, we worked with two types of PET coated ECCS, denoted by PET039 and PET052. Two copies were made of each deformation state, resulting in 8 samples.

We have studied the height of the defects in the can wall and the deformation of the defects as a function of the height and compared the results with the geometrical model.

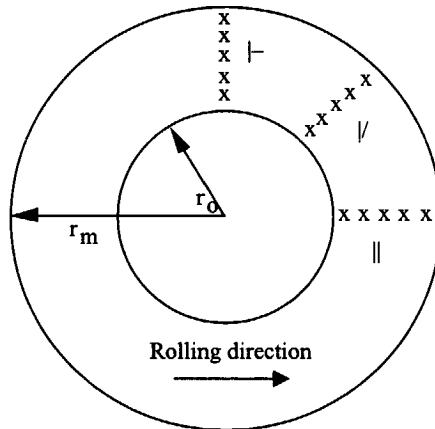


Figure 4.2. Location of round defects in the initial plane sheet. The positions of the defects are marked by the symbol x.

4.3.1 Height of the defects in the cups

In table 4-I, the deformation parameters of the different cups are shown. The different deformation steps are denoted by "1xD" en "2xD" after 1 and 2 deep drawing steps and by "1xW" and "3xW" after 1 and 3 rings of wall ironing. Also the wall thickness and the total height of the cups are given. The wall thickness was measured by a micrometer and did not vary as a function of the height, confirming the assumption of section 4.2.

Due to the so-called "earing-effect" the total height of the cup was not well defined within ± 1 mm, and the values in table 4-I are averages. Also the theoretically expected cup heights according to equation (4.3) are shown in this table, and we see that these values are more or less in agreement with the measurements.

	1xD	2xD	1xW	3xW
a (μm)	325	325	325	325
b (μm)	325	325	255	150
r_o (cm)	5.00	3.75	3.75	3.75
r_m (cm)	7.45	7.45	7.45	7.45
z_m (cm), th.	3.05	5.53	7.04	11.97
z_m (cm), ms.	3.15	5.1	7.2	11.5

Table 4-I. Deformation parameters of the different cups that were studied and the total height z_m of the cups. The abbreviations "th." and "ms." denote from theory or from measurement.

R (cm)	1xD	2xD	1xW	3xW
4.65	- *	1.0	1.1	1.7
5.20	- **	1.7	2.1	3.2
5.75	0.8	2.5	3.1	5.0
6.30	- **	3.4	4.2	6.9
6.85	2.2	4.3	5.5	9.0

Table 4-II. Measured heights of the artificial defects. * Artificial defect in the bottom of the can, ** Not available.

Assuming that the artificial defects in the PET coating follow the underlying steel, one can determine the movement of the volume-elements during the DWI-process by measurement of the position of the defects after the different deformation steps. In table 4-II, the measured heights of the defects are given. In figure 4.3, these values are plotted versus the theoretically expected heights according to equation (4.2).

We see that after the deep drawing steps, the height of the defects is in good agreement with the expected values. However, after wall ironing significant deviations are visible. After the first ring, the whole series of defects was found about 1mm lower than expected and after 3 rings, they were even about 5mm lower. This could be explained by the fact that a profile is created in the bottom of the can during wall ironing, by which the material in the can wall is drawn downwards.

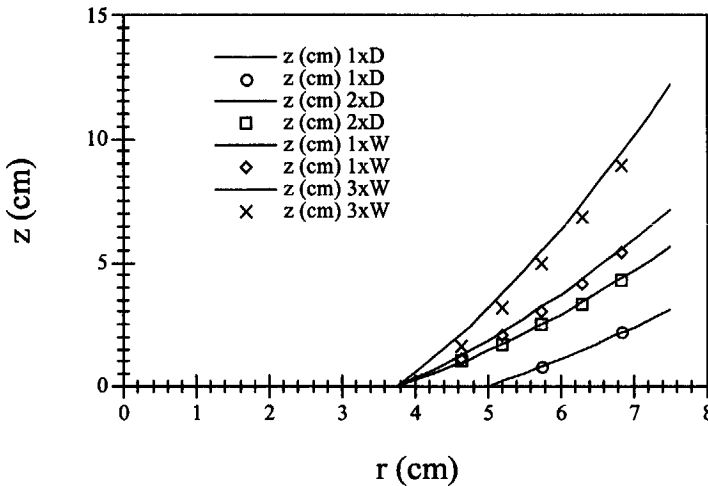


Figure 4.3. Measured heights of the artificial defects after the different can formation steps. The solid lines are the theoretically expected values.

We observed no influence of the rolling direction on the height of the defects. Also the type of PET (039 or 052) had no influence and the defects on the inside and on the outside of the can were all found at the same heights.

An interesting and unexpected observation is that during wall ironing most of the defects had disappeared. In fact, after 1 ring of wall ironing, the defects inside the can were still visible, but from the outside defects, the 3 highest defects had disappeared. And after 3 rings of wall ironing, all defects inside and outside the can wall were closed, although white stripes of about 1cm long were visible at the location where the defects were expected. That the defects were really closed was verified by the CuSO_4 penetrant test, mentioned in chapter 3. In fact, the disappearance of the defects could have made it tricky to relocate them, but prior to making the defects, the desired positions had been marked by a permanent overhead marker pen. After the deformation steps, these original marks had faded away, but were still visible and allowed us to determine the positions of the defects with reasonable accuracy.

4.3.2 Deformation of the defects in the cups

During deep drawing and wall ironing, the originally round defects are deformed to ellipses. The long axis of the ellipses was always parallel to the axis of the can, which was also expected from the geometrical model of section 4.2. The ratio of the long and short axis of the ellipses was measured by a Vickers microscope and was taken as a measure for the overall deformation of the defects. This ratio is equal to the ratio of the axial and azimuthal deformations ϵ_z and ϵ_θ . The predictions from the geometrical model for this ratio follow from equations (4.4) and (4.5):

$$\frac{\epsilon_z}{\epsilon_\theta} = \frac{a}{b} \left(\frac{r}{r_0} \right)^2 \tag{4.7}$$

This quantity is plotted in figure 4.4 for the different can formation steps. The ratio $\epsilon_z / \epsilon_\theta$ is always larger or equal than 1, as follows from the fact that $a \geq b$ and $r \geq r_0$.

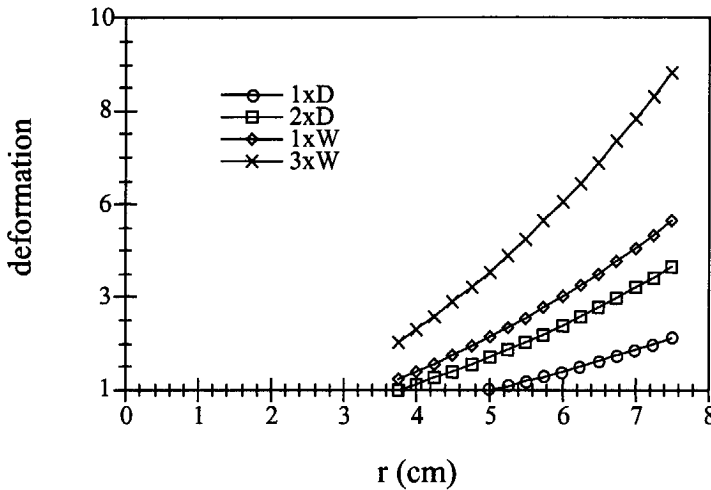


Figure 4.4. Theoretical deformation for the different can formation steps. The plotted deformation is equal to the ratio $\epsilon_z / \epsilon_\theta$.

In the figures 4.5a-f and 4.6a-f, the measured deformations after two steps of deep drawing and after 1 ring of wall ironing are shown. Not shown are the data after the first deep drawing step, which were not available and after 3 rings of wall ironing, in which case the defects were closed as discussed in section 4.3.1. It follows that:

- All defects are stretched in the axial direction;
- Higher in the can wall, the deformation is larger.

However, the measured deformation of the defects is larger than expected from the geometrical model. This could in fact also be said for the defects after 3 rings of wall ironing, in which case the defects are so deformed that they are closed.

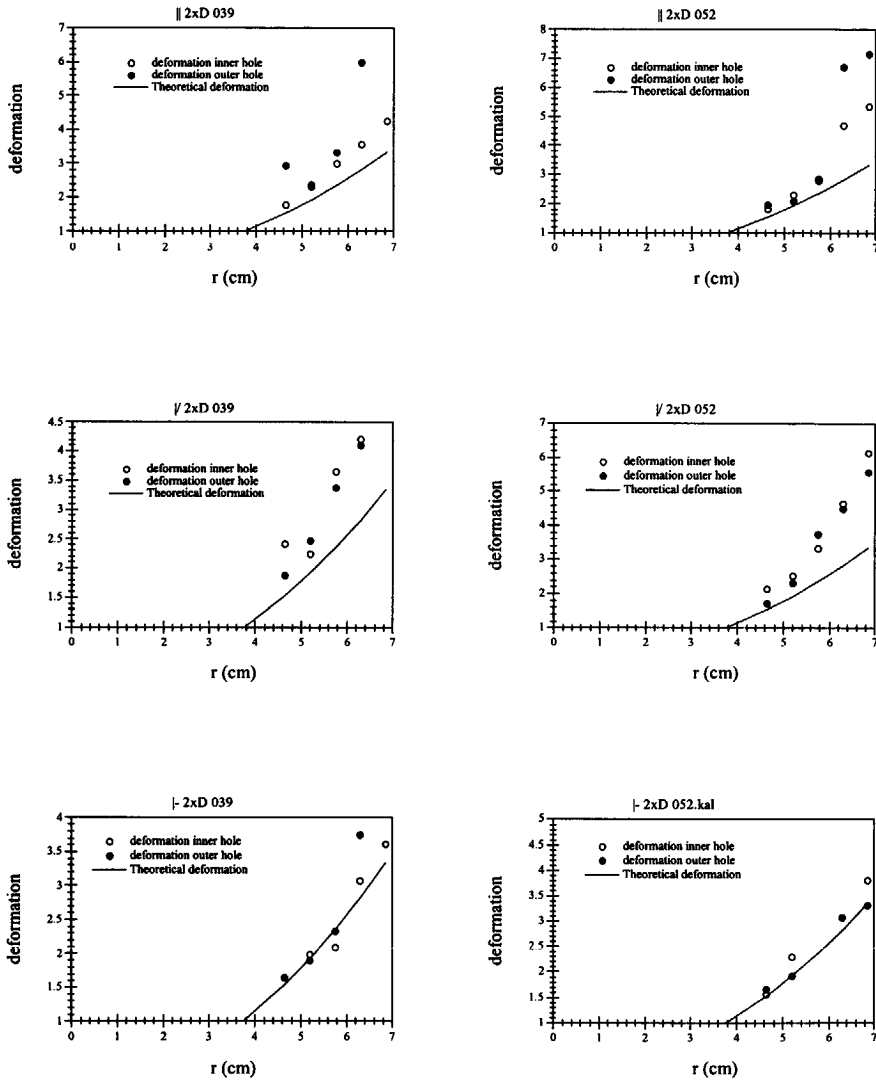


Figure 4.5a-f. Deformation of the artificial defects after 2 deep drawing steps. The solid lines are the expected deformations from the geometrical model, according to equation (4.7).

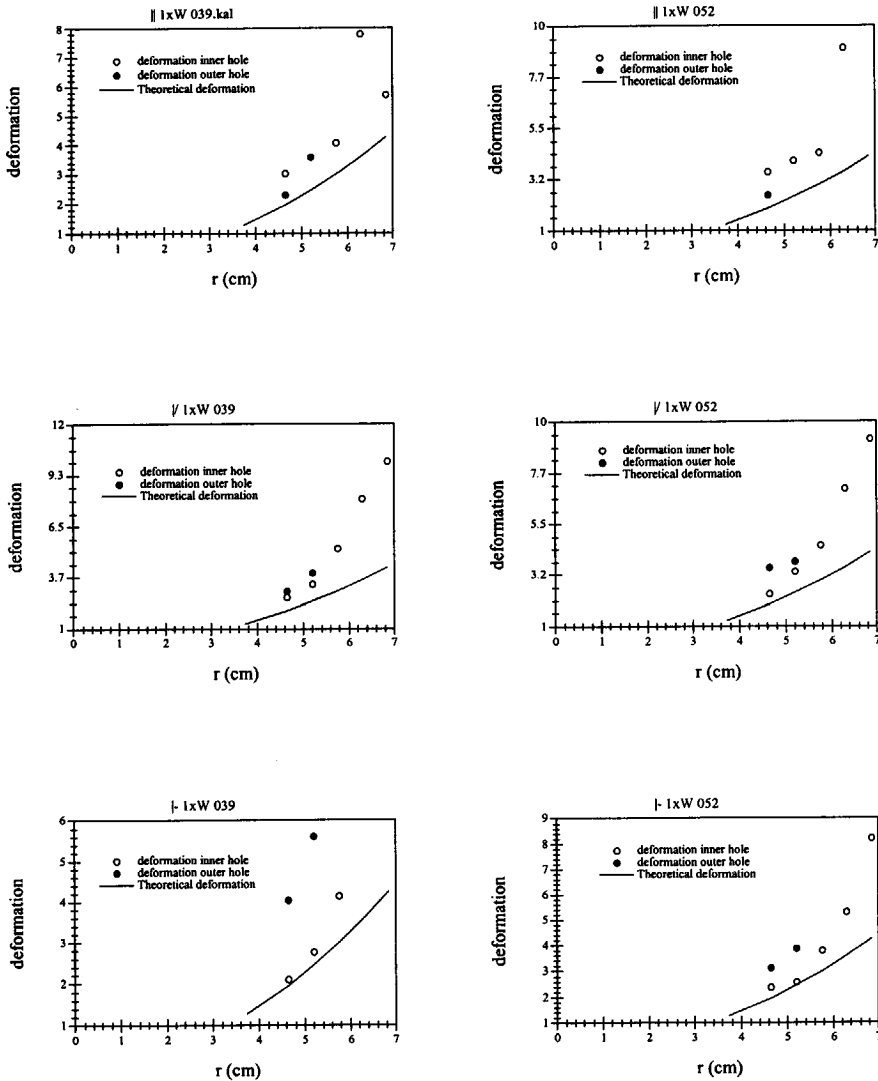


Figure 4.6a-f. Deformation of the artificial defects after 1 ring of wall ironing. The solid lines are the expected deformations from the geometrical model, according to equation (4.7).

Due to statistical errors, it is difficult to draw any conclusions on the influence of the PET type on the deformation of the defects. The same holds for the differences between the defects inside and outside the cups.

The rolling direction seems to have some influence on the deformation of the defects, at least for the deep drawing process, as may be concluded from figures 4.5a-f. This can be understood from the fact that the ridges of the rolling process, which are directed parallel to the rolling direction, have some grip on the coating. Thus one might expect that the deformation is less for deep drawing and wall ironing parallel to the rolling direction. This is indeed the case after two deep drawing steps, as follows from figure 4.5, but after wall ironing the effect is no longer visible according to figure 4.6.

4.4 Summary and discussion

We have studied the position and deformation of artificial defects in PET coated ECCS when the PET coated ECCS was deformed into a can by a DWI-process. The defects were initially round $70\mu\text{m}$ holes in the PET coating. After the DWI-process we observed the following:

- The defects were axially stretched and azimuthally compressed;
- The deformation was larger after each following step of the DWI-process;
- This deformation was larger in the top of the can than in the bottom;
- The defects had disappeared after 3 rings of wall ironing.

It was expected that the deformation of the defects would be a measure for the local deformation of the coating. However, the deformation of the defects was larger than expected from a simple geometrical model. This is probably due to the fact that under the high pressures of the DWI-process, the coating flows and is pushed into the defects. Eventually after the last step of the DWI-process the defects had even disappeared. This is very interesting because it means that the DWI-process has a sort of "curing" effect on small defects in the base material.

But although the measurements and the simple geometrical model do not completely agree on the degree of deformation, it is clear that the coating is asymmetrically deformed and that the deformation is higher in the top of the can. This may seem obvious, but was previously denied by several researchers.

Something else that was thought impossible was that marks of transparency pens on the coating would survive the DWI-process. Nevertheless we found that after two deep drawing steps, such marks were still clearly visible and even after one or even three rings of wall ironing the marks were easily retraceable although they had faded. This means that instead of artificial defects, one can also use transparency pens for studying the displacement and deformation of the coating. (In fact we did this to determine the location of the defects that were closed by the DWI-process.) Studying the deformation of the coating by marks of transparency pens is perhaps simpler than by artificial defects and may even lead to more accurate results since the coating flows into the defects at high pressure. This disturbing effect does not play a role for the marks.

We studied the deformation of two different types of PET (039 and 052) on ECCS, but we did not see significant differences due to random errors. For the same reason, we also did not see a significant difference between the deformation inside or outside the can. We also studied the influence of the rolling direction on the deformation of the defects, but did not observe much difference, apart from an effect after the deep drawing steps. In that case, the deformation of defects was less at locations where the base material was stretched perpendicularly to the rolling direction.

The random errors in the measurements will be due to inhomogeneities in the PET coatings. Especially in PET039, many air bubbles were observed. In the base material, these bubbles were round and one could think of using the bubbles to ones advantage by using their shape as a measure of deformation during the DWI-process.

The experiments of this chapter have shown that after the DWI-process, the deformation of the coating is different at different locations in the coating. Therefore in future studies of the corrosion protection, such as by the methods of chapter 2 and 3 (or perhaps by the method of chapter 6), different parts of the can have to be considered.



Chapter 5

The application of electric potential scans for studying localized corrosion

5.1 Introduction

With standard methods for corrosion research such as polarization measurements and electrochemical impedance spectroscopy (EIS), one determines the corrosion current, which is a measure of the corrosion rate. With these methods a metal surface is covered by an electrolyte layer, an external voltage is applied between the metal and the electrolyte, and the resulting current through the interface is measured as a function of the applied voltage. Two important limitations of these techniques are:

- The corrosion current of the whole covered area is measured;
- A reference electrode has to be put in the electrolyte.

The first requirement means that one cannot see lateral differences of the corrosion rate and therefore, one cannot study localized corrosion processes. This is a serious drawback, because many practical corrosion processes are forms of localized corrosion. The second limitation means that one cannot study two important forms of corrosion, which are corrosion under intact coatings and atmospheric corrosion. (In the latter case, the electrolyte layer is too thin to insert an electrode.)

Recently, a number of new techniques have been introduced for localized corrosion studies. With these techniques, the electric potential above a corroding surface is scanned by a microelectrode. Isaacs introduced the Scanning Reference Electrode Technique (SRET) and the Scanning Vibrating Electrode Technique (SVET) [Isaacs 1980, 1988] and Stratmann introduced the Scanning Kelvin Probe (SKP) [Stratmann 1987, 1990]. The main difference between these techniques is that with the SRET and the SVET, the electric potential in the electrolyte is measured while the SKP scans the electric potential above the electrolyte. Therefore the SRET and the SVET are applied to forms of corrosion where the substrate is immersed in a relatively thick electrolyte while the SKP is rather used for studying atmospheric and underpaint corrosion.

The SRET is in fact a reintroduction of a very old technique for experimentally solving the Laplace equation in 2 or 3 dimensions. With this technique, the desired geometry was built inside a tank filled with electrolyte, electric potentials were applied on the electrodes and the electric potential in the electrolyte was scanned by a microelectrode (cf. [Makar 1948, Richards 1964]). Interesting techniques have been suggested to determine the electric field more rapidly than by a scanning probe (cf. [Burt 1957]), but the “electrolytic plotting tank” has fallen into disuse since the introduction of electronic computers. The principle of the SVET was also known previously and for instance used to study the role of electric fields in cell growth processes [Jaffe 1974]. The SKP is already used for many years for work function differences between metals [Kelvin 1898, Zisman 1932].

In this chapter, we will not discuss the experimental details of the electric potential scanners, but assume that they can perfectly measure the electric field in or above the electrolyte on a corroding surface. The question posed here is rather what can be learned from an electric potential scan above a corroding surface in terms of parameters of the corrosion process. We will therefore study the influence of corrosion processes on the electric potential above a corroding surface, as well as a number of phenomena other than corrosion, which also influence the electric potential above a surface.

5.2 The electric field above a corroding surface

First in section 5.2.1, we start by a discussion on the relatively simple case of clean metal surfaces without an electrolyte layer. This is a useful introduction to the case of metals with an electrolyte layer, which is discussed in section 5.2.2. In these first two sections, a “bulk treatment” is given, which is sufficient for the main discussion on the application of electric potential scans for localized corrosion studies. The case of coated metals brings additional difficulties and is covered by the discussions in sections 5.2.3 and 5.2.4.

5.2.1 Metals

One might expect the electric potential to be constant on a clean and perfectly conducting metal surface, but this will generally not be the case when the surface is an inhomogeneous mixture of different types of metals. Then the electrons are not equally “comfortable” at different positions on the surface and will accumulate in the preferred areas. As a result, these areas become negatively charged while the less preferred areas become positively charged. This inhomogeneous distribution of charge results in an inhomogeneous electric surface potential.

A convenient way to describe the electron distribution in a metal, or on a metal surface, is to consider the chemical potential of the electrons. In general, the chemical potential of a particle is the “internal chemical potential” plus the sum of all possible potential energy contributions to the particle (cf. [Kittel 1980]). Hence for a particular type of particle, denoted by j , one has:

$$\mu_j = k_B T \ln(n_j / n_j^0) + q_j V + M_j g h + \dots \quad (5.1)$$

The first term, which is the internal chemical potential, contains the particle density n_j , the temperature T , the Boltzmann constant $k_B = 1.4 \times 10^{-23} \text{ J/K}$ and the quantum concentration n_j^Q , which is a reference concentration. The second term is the electric potential energy, where q_j is the charge per particle and V is the electric potential and the third term is the potential energy of gravitation. Here M_j is the mass per particle, $g = 9.81 \text{ m/s}^2$ and h is the altitude.

The chemical potential describes the tendency of particle movement. In fact, the force on a particle is minus the gradient of the chemical potential. Therefore, particles tend to move from higher to lower chemical potential values. For the gravitation term for example, this results in the gravitational force towards lower altitudes. The electric term on the other hand leads to a force towards lower electric potentials for positively charged particles and towards higher electric potentials for negatively charged particles. The internal chemical potential finally pushes the particles to lower concentrations. This process is nothing but diffusion.

As long as the particles experience a net force, there will be a net particle movement and their spatial distribution will change. This goes on until the chemical potential has become the same for all particles. Then the net force on the particles is zero and the thermodynamic equilibrium state is reached. It follows that the condition:

$$\mu_j(r) = \text{constant} \tag{5.2}$$

describes the spatial particle distribution in thermodynamic equilibrium. Here, r is the spatial coordinate. For example, the equilibrium distribution of electrically neutral particles in the earth's atmosphere:

$$n_j(h) = n_j(0) \exp[-M_j g h / k_B T]$$

follows immediately from (5.1) and (5.2), with $r = h$. Here, $n_j(0)$ is the particle density at ground level and it is assumed that the temperature is independent of height.

The spatial distribution of electrons in a metal in thermodynamic equilibrium follows from similar considerations, but there is a complication: the expression for the internal chemical potential in (5.1) is only valid for concentrations below the quantum concentration. This holds for many cases, such as gases in a gas mixture at atmospheric pressures or for ions in dilute solutions but not for electrons in metals that have to be regarded as a quantum gas (cf. [Kittel 1980]).

However, for the spatial distribution of electrons in metals a good approximation can be obtained by considering the internal chemical potential as independent of the electron concentration. Thus we write the following for the chemical potential of the electrons in a metal (ignoring the gravitational contribution here and below):

$$\mu_e = \mu_e^{\text{int}} + q_e V \tag{5.3}$$

This can be done because for very high concentrations of charged particles, the influence of the particle concentration on the electric potential V is much more important than the influence on μ_e^{int} . This is in fact not so easily shown for a quantum gas, but is quite obvious for classical particles since V changes linearly with n_j while μ_e^{int} is proportional to $\ln(n_j)$. Now from (5.3) and the thermodynamic equilibrium condition (5.2) it follows that:

$$V(r) = \text{constant} - \mu_e^{\text{int}}(r)/q_e$$

Hence in thermodynamic equilibrium, the electrons will be distributed over the metal such that V has the same shape as the internal chemical potential μ_e^{int} . This is illustrated by figure 5.1. Note that the electron charge is negative ($q_e = -1.6 \times 10^{-19} \text{C}$), otherwise $V(r)$ and μ_e^{int} would have opposite shape.

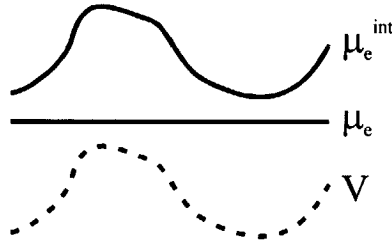


Figure 5.1. *The spatial distribution of the internal chemical potential, the chemical potential and the electric potential for electrons in thermodynamic equilibrium.*

It follows that in a homogeneous metal, where μ_e^{int} is constant, the electric potential V is constant. Hence a homogeneous metal surface is indeed an equipotential surface. On an inhomogeneous metal surface however, such as on the surface of an imperfect alloy, μ_e^{int} and hence V will spatially vary.

Inhomogeneities on a metal surface can also be abrupt, such as at the contact between two different metals. Then the potential will make a discrete step when crossing the interface between the metals and the magnitude of this step reads:

$$V_1 - V_2 = -(\mu_e^{\text{int}1} - \mu_e^{\text{int}2})/q_e$$

where $\mu_e^{\text{int}1}$ and $\mu_e^{\text{int}2}$ are the internal chemical potentials of the different metals. The magnitude of the potential difference between two metals in electric contact is known as the "contact potential difference" (CPD) and can be in the order of a volt (cf. [Weast 1975]). In many practical situations CPDs occur, for instance in electric circuits where different metals are used. A well-known application is found in thermocouples that use the temperature dependence of the CPD for temperature measurements. The CPD is illustrated by figure 5.2 for two metals for which $\mu_e^{\text{int}2} > \mu_e^{\text{int}1}$. After the metals are electrically connected, some electrons travel to metal 1 with the lower chemical potential, which becomes negatively charged. This lowers the electric potential of metal 1 and (since q_e is negative) raises the chemical potential until in thermodynamic equilibrium, μ_e is the same in metal 1 and 2.

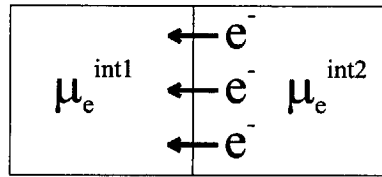


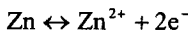
Figure 5.2. Two metals with different internal chemical potentials ($\mu_e^{int2} > \mu_e^{int1}$), just after the moment that they are electrically connected.

5.2.2 Electrolytes on metals

The electric potential in electrolyte layers on metals can be described by the same methods as in the previous section in case the electrolyte layer is in thermodynamic equilibrium with the metal, which is the first case discussed. When corrosion processes occur, the dynamics of the chemical reactions must also be considered. This will first be done for a uniformly corroding surface, and then for localized corrosion processes. Special attention is paid to the relation between the electric potential in the electrolyte and the corrosion rates, with the application of electric potential scanners for corrosion research in mind.

Thermodynamic equilibrium

Similar to the CPD between metals, the origin of the voltage over a metal-electrolyte interface is a charged particle equilibrium. But instead of electrons, the charged particles are ions. For instance, at the interface of a zinc and water, the Zn^{2+} ion can cross the interface by the reaction:



Now the equilibrium of the zinc ions can be described in exactly the same way as the electron equilibrium between two different metals, that is by considering the chemical potentials of the zinc ions on both sides of the interface. In the metal we have:

$$\mu_{Zn^{2+}}(m) = \mu_{Zn^{2+}}^{int}(m) + 2|q_e|V_m \tag{5.4}$$

where V_m is the potential of the solid zinc and $2|q_e|$ is the charge of the zinc ions. The internal chemical potential of the zinc ions in solid zinc can be considered constant since the concentration of zinc in solid zinc is determined by the crystal structure. In the electrolyte we have:

$$\mu_{Zn^{2+}}(e) = k_B T \ln(n_{Zn^{2+}} / n_{Zn^{2+}}^0) + 2|q_e|V_e$$

In this case, the concentration dependent term is significant, since the concentration of Zn^{2+} can be changed by several orders of magnitude by for instance dissolving zinc sulfate in the

electrolyte. Instead of the quantum concentration, the more usual reference concentration for ions in solutions is 1 mol per liter. Then:

$$\mu_{\text{Zn}^{2+}}(e) = \mu_{\text{Zn}^{2+}}^{\text{ref}}(e) + k_B T \ln(n_{\text{Zn}^{2+}} / n_{\text{Zn}^{2+}}^{\text{ref}}) + 2|q_e|V_e \quad (5.5)$$

Often the reference concentration is left out of the expression, which fixes the unit of the ion concentration in moles per liter. This however also violates the logical form of the expressions and will not be done here. For the same reason, the usual replacement of the natural logarithm by a common logarithm will not be done, nor will we substitute actual numbers for $k_B T$ and the electron charge q_e .

Now in thermodynamic equilibrium, the chemical potentials of the zinc ions given by (5.4) and (5.5) must be equal, which leads to the following:

$$V_m - V_e = (V_m - V_e)^{\text{ref}} + \frac{k_B T}{2|q_e|} \ln \left[\frac{n_{\text{Zn}^{2+}}}{n_{\text{Zn}^{2+}}^{\text{ref}}} \right]$$

This formula, first written by Nernst, describes the voltage over a metal-electrolyte interface as a function of the concentration of metal ions in the electrolyte [Nernst 1889]. Here $(V_m - V_e)^{\text{ref}}$ is the voltage at the reference concentration 1 mole of zinc ions per liter of electrolyte. We see that $V_m - V_e$ increases with the Zn^{2+} concentration in the electrolyte. This can be understood from the fact that for low Zn^{2+} concentrations, there will be a tendency to dissolve and for high concentrations, there will be a tendency to precipitate. In thermodynamic equilibrium, the electrolyte will therefore be more positive ($V_e > V_m$) for very low Zn^{2+} concentrations and more negative ($V_e < V_m$) for very high Zn^{2+} concentrations.

For electrolytes with different kinds of particles, the chemical potential μ_j of each species is constant in thermodynamic equilibrium. When the particles can mutually react with each other and with the metal electrode, the situation is more complicated, because the numbers of particles N_j of the species are not fixed and are also parameters. A convenient way to describe the system is then to consider the Gibbs free energy (cf. [Kittel 1980]):

$$G = \sum_j N_j \mu_j$$

which can be regarded as a kind of overall chemical potential. Now for systems in thermal and pressure equilibrium with a large reservoir, the Gibbs free energy G is conserved. (This large reservoir is often the earth's atmosphere.) This leads to the following condition:

$$dG = \sum_j \mu_j dN_j = 0 \quad (5.6)$$

Hence, a virtual variation of the particle concentrations around the equilibrium values at constant temperature T and pressure p leaves G constant in first order. (At this variation, the chemical potentials μ_j are constant because these are functions of the intrinsic system variables T and p .) Now suppose that the following chemical reaction can take place between the different species A_j :

$$\sum_j \nu_j A_j \leftrightarrow 0 \quad (5.7)$$

Here, ν_j is the reaction coefficient of a species. These are not unique. For instance, for the hydrogen evolution reaction $2\text{H}^+ + 2\text{e}^- \leftrightarrow \text{H}_2$ one can choose $\nu(\text{H}^+) = 2$, $\nu(\text{e}^-) = 2$ and $\nu(\text{H}_2) = -1$, but one could also choose multiples of these numbers. Now by substitution of expressions such as (5.3), (5.4) and (5.5) for the relevant chemical potentials in (5.6), the following can be obtained:

$$V_m - V_e = (V_m - V_e)^{\text{ref}} + \frac{k_B T}{\nu_e |q_e|} \ln \left[\prod_{j \neq e} \left(\frac{n_j}{n_j^{\text{ref}}} \right)^{\nu_j} \right] \quad (5.8)$$

This is the voltage over the metal-electrolyte interface as a function of the particle concentrations for the equilibrium of the general reaction (5.7). In appendix A, the derivation of (5.8) is carried out in greater detail.

Corrosion

When corrosion reactions take place at a metal-electrolyte interface, the system is not in equilibrium and the voltage over the interface will differ from the equilibrium voltage $\Delta V^{\text{eq}} = V_m - V_e$, given by (5.8). How much the actual ΔV differs from ΔV^{eq} depends on the reaction kinetics and is the main subject of the discussion below.

As an example, we will consider the corrosion of an iron electrode in diluted sulfuric acid, in which case the following electrochemical reactions can take place:



These reactions will be referred to as the anodic and cathodic reactions of the corrosion process. Each of these reactions has an equilibrium voltage, which will be denoted by ΔV_a^{eq} and ΔV_c^{eq} . Now when ΔV_a^{eq} and ΔV_c^{eq} are not the same, the system cannot be in equilibrium, since ΔV over the metal-electrolyte interface must have a definite value. This value will lie somewhere in between ΔV_a^{eq} and ΔV_c^{eq} and it now follows from the relative magnitudes of ΔV_a^{eq} and ΔV_c^{eq} in which direction the reactions (5.9a) and (5.9b) will take place.

In case $\Delta V_a^{\text{eq}} > \Delta V_c^{\text{eq}}$, the actual voltage ΔV will be smaller than ΔV_a^{eq} . Then the metal is more negative and the electrolyte is more positive than in equilibrium and the iron ions will tend to precipitate. Then the iron electrode will not corrode. On the other hand, when $\Delta V_a^{\text{eq}} < \Delta V_c^{\text{eq}}$, it is the opposite. Then the actual ΔV will be larger than ΔV_a^{eq} , the metal ions will tend to dissolve and the iron electrode does corrode. (The same conclusions can of course be found by considering the hydrogen dissolution reaction.)

The relative magnitude of ΔV_a^{eq} and ΔV_c^{eq} depends on the particle concentrations, but normally ΔV_a^{eq} will be smaller than ΔV_c^{eq} and the iron will corrode. (In fact by definition, the anodic reaction is the reaction with the lower ΔV^{eq} .) For example, when the concentrations of Fe^{2+} and H^+ ions are equal to the standard reference concentrations of 1 mol per liter and

when the electrolyte is in equilibrium with hydrogen gas with a partial pressure of 1 atm, ΔV_a^{eq} is about 0.44V lower than ΔV_c^{eq} (cf. Weast 1975).

Now when the corrosion reaction takes place, the electric currents of the Fe^{2+} and H^+ ions at the interface must balance, otherwise the metal and the electrolyte would not remain electrically neutral. Denoting these currents by the anodic and cathodic currents I_a and I_c respectively, we have:

$$I_a + I_c = 0 \quad (5.10)$$

In the more general case, when more than two reactions take place, the sum of the currents must be zero. The anodic current is a measure for the corrosion rate of the iron electrode and is therefore a very important quantity for corrosion studies. The magnitude of I_a follows from how much the voltage ΔV over the metal-electrolyte interface differs from ΔV_a^{eq} . A general relation between the electric current I_k and the deviation $(\Delta V - \Delta V_k^{eq})$ from the equilibrium voltage of a chemical reaction is given by the Butler-Volmer equation [Butler 1924, Erday-Grúz 1930]:

$$I_k = I_0^k \left[\exp\left(\frac{\alpha_k q_k (\Delta V - \Delta V_k^{eq})}{k_B T}\right) - \exp\left(-\frac{(1 - \alpha_k) q_k (\Delta V - \Delta V_k^{eq})}{k_B T}\right) \right] \quad (5.11)$$

This equation holds for particles with a charge q_k , crossing a potential barrier, which is characterized by the exchange current I_0^k and the parameter α_k . When $\Delta V = \Delta V_k^{eq}$, the reaction is in equilibrium and the net current through the interface I_k is zero. The exchange current I_0^k is a measure for the speed of the reaction, since for large I_0^k , I_k changes relative rapidly as ΔV differs more from ΔV_k^{eq} . The parameter α_k is a positive number between 0 and 1 describing the "asymmetry" of the potential barrier (for a symmetrical potential barrier, $\alpha_k = 1/2$). More details can be found in appendix B, where a discussion on the Butler-Volmer equation is given.

Both for I_a and I_c , an equation of the form (5.11) can be written and with $I_a = -I_c$, according to (5.10), the actual voltage ΔV over the metal-electrolyte interface can be calculated. A useful approximation of the solution can be obtained from considering the asymptotic behavior of $I_a(\Delta V)$ and $I_c(\Delta V)$ for $\Delta V \gg \Delta V_a^{eq}$ and $\Delta V \ll \Delta V_c^{eq}$:

$$I_a \cong I_0^a \exp\left[\frac{\alpha_a q_a (\Delta V - \Delta V_a^{eq})}{k_B T}\right] \cong I_0^a \exp\left[\frac{(\Delta V - \Delta V_a^{eq})}{b_a}\right]$$

$$I_c \cong -I_0^c \exp\left[-\frac{(1 - \alpha_c) q_c (\Delta V - \Delta V_c^{eq})}{k_B T}\right] \cong -I_0^c \exp\left[-\frac{(\Delta V - \Delta V_c^{eq})}{b_c}\right]$$

Here the parameters b_a and b_c are defined, which are known as the anodic and cathodic Tafel slopes after Julius Tafel, who experimentally observed the asymptotic behavior prior to the formulation of the Butler-Volmer equations [Tafel 1900]. With these approximations and with $I_a = -I_c$, the following approximation for the voltage over the metal-electrolyte interface can be obtained:

$$\Delta V \equiv \Delta V_{\text{corr}} \cong \frac{b_c}{b_a + b_c} \Delta V_a^{\text{eq}} + \frac{b_a}{b_a + b_c} \Delta V_c^{\text{eq}} + \frac{b_a b_c}{b_a + b_c} \ln \left(\frac{I_0^c}{I_0^a} \right) \quad (5.12)$$

This result is given by Stern and Geary [Stern 1957]. The actual voltage over metal-electrolyte interface during corrosion is called the corrosion voltage and denoted by ΔV_{corr} . (This voltage is also known as the mixed voltage, which is perhaps indeed a better name. In literature, one more often encounters the “corrosion potential” or the “mixed potential”. However, in our case this would be inappropriate since in our discussion, ΔV is a well-defined potential difference and not a potential with respect to an some reference potential.)

According to (5.12), ΔV_{corr} is a sort of weighed average of ΔV_a^{eq} and ΔV_c^{eq} . As mentioned above, ΔV_{corr} must lie between ΔV_a^{eq} and ΔV_c^{eq} . The last term in the latter expression seems to contradict that, but when this last term is dominant, the Stern-Geary approximation is not valid. The exact solution for ΔV_{corr} is illustrated by figure 5.3, where the anodic and cathodic currents I_a and I_c are drawn separately as a function of ΔV . When I_a is equal to $-I_c$, ΔV is equal to ΔV_{corr} . The values of I_a or I_c also follow from the Stern-Geary approximation. We will give the result for the anodic current, which is proportional to the corrosion rate and therefore also called the “corrosion current”:

$$I_{\text{corr}} \equiv I_a(\Delta V_{\text{corr}}) = I_0^a \exp \left[\frac{\Delta V_{\text{eq}}^c - \Delta V_{\text{eq}}^a}{b_a + b_c} + \frac{b_c}{b_a + b_c} \ln \left(\frac{I_0^c}{I_0^a} \right) \right] \quad (5.13)$$

This must always result in a positive value for I_{corr} ; otherwise the Stern-Geary approximation is not valid.

Now we come back to the original question whether the deviation of the electric potential from thermodynamic equilibrium can be regarded as a measure for the corrosion rate. This question is equivalent to the question how I_{corr} is related to ΔV_{corr} . However, by equations (5.12) and (5.13), one can show that an increased I_{corr} can both correspond to an increased or to a decreased ΔV_{corr} . This can already be shown by considering the influence of the H^+ and Fe^{2+} concentrations. First we observe from (5.13) that I_{corr} is increased when either:

- The concentration of H^+ ions is increased, leading to a more positive ΔV_c^{eq} ;
- The concentration of Fe^{2+} ions is diminished, leading to a more negative ΔV_a^{eq} .

Besides it follows from equation (5.12) that in the first case, the corrosion voltage ΔV_{corr} is increased while in the second case, ΔV_{corr} is decreased. But in both cases, the corrosion rate is increased, hence the corrosion voltage alone is no measure for the corrosion rate. (By a combination of the cases, it may even be that the corrosion voltage is the same for a higher corrosion rate.) In a way this might be expected in advance, because I_{corr} and ΔV_{corr} are determined by too many thermodynamic and kinetic parameters. A one on one relation between I_{corr} and the deviations of the potential from thermodynamic equilibrium (hence ΔV_{corr}), is therefore not possible.

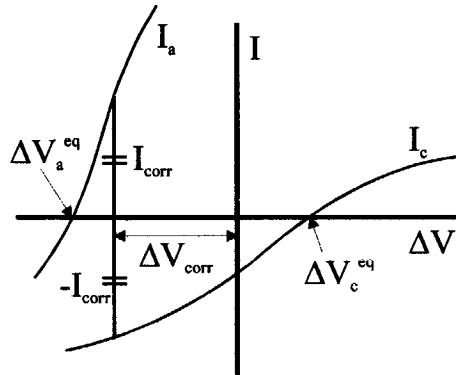


Figure 5.3. The corrosion voltage and the corrosion current for given anodic and cathodic Butler-Volmer curves. Here, the voltage $\Delta V = V_m - V_e$ is the potential difference between the metal and the electrolyte.

Localized corrosion

So far we considered uniform corrosion, where the corrosion rate is uniformly distributed over the metal surface. Then the anodic and cathodic reactions take place everywhere on the surface, with uniformly distributed anodic and cathodic currents.

With localized corrosion, the anodic and cathodic reactions are spatially separated over anodic and cathodic areas. This separation always occurs due to some kind of asymmetry on the surface. This may for example be a variation of the metal composition on the surface, such for instance a copper patch on an iron surface. When such a surface is covered by diluted sulfuric acid, the cathodic reaction takes place on the copper site and the anodic reaction takes place on the iron. The anodic and cathodic reactions are then spatially separated such as in a battery as depicted in figure 5.4. It is essential that the copper patch makes electric contact with the iron, otherwise the battery circuit is not closed.

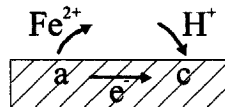
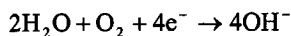


Figure 5.4. Electric circuit between the anodic sites (a) and cathodic sites (c). The electric current is carried by electrons in the metal and by ions in the electrolyte.

Another common cause of localized corrosion is a different aeration of different parts of an iron surface. This can for instance occur when the surface is partly covered with dirt or when the electrolyte layer is locally thinner. At the well-aerated areas, the electrolyte is more easily supplied by oxygen, and the following reaction will occur:



Under normal conditions, the equilibrium voltage of this reaction is considerably more positive than for the iron dissolution reaction. Therefore the well-aerated areas of an iron surface will be cathodic, and the poorly aerated areas will be anodic.

Now let us consider the electric potential in the electrolyte V_e in case of a localized corrosion process. The easiest case is when the corrosion current between the anodic and cathodic sites is very small. Then the anodic and cathodic sites are almost in equilibrium and at the anodic site, $\Delta V_a = (V_m - V_e)_a = \Delta V_a^{\text{eq}}$ while at the cathodic site, $\Delta V_c = (V_m - V_e)_c = \Delta V_c^{\text{eq}}$. Since $\Delta V_a^{\text{eq}} < \Delta V_c^{\text{eq}}$ (otherwise the system does not corrode), it follows that V_e is higher at the anodic site and lower at the cathodic site.

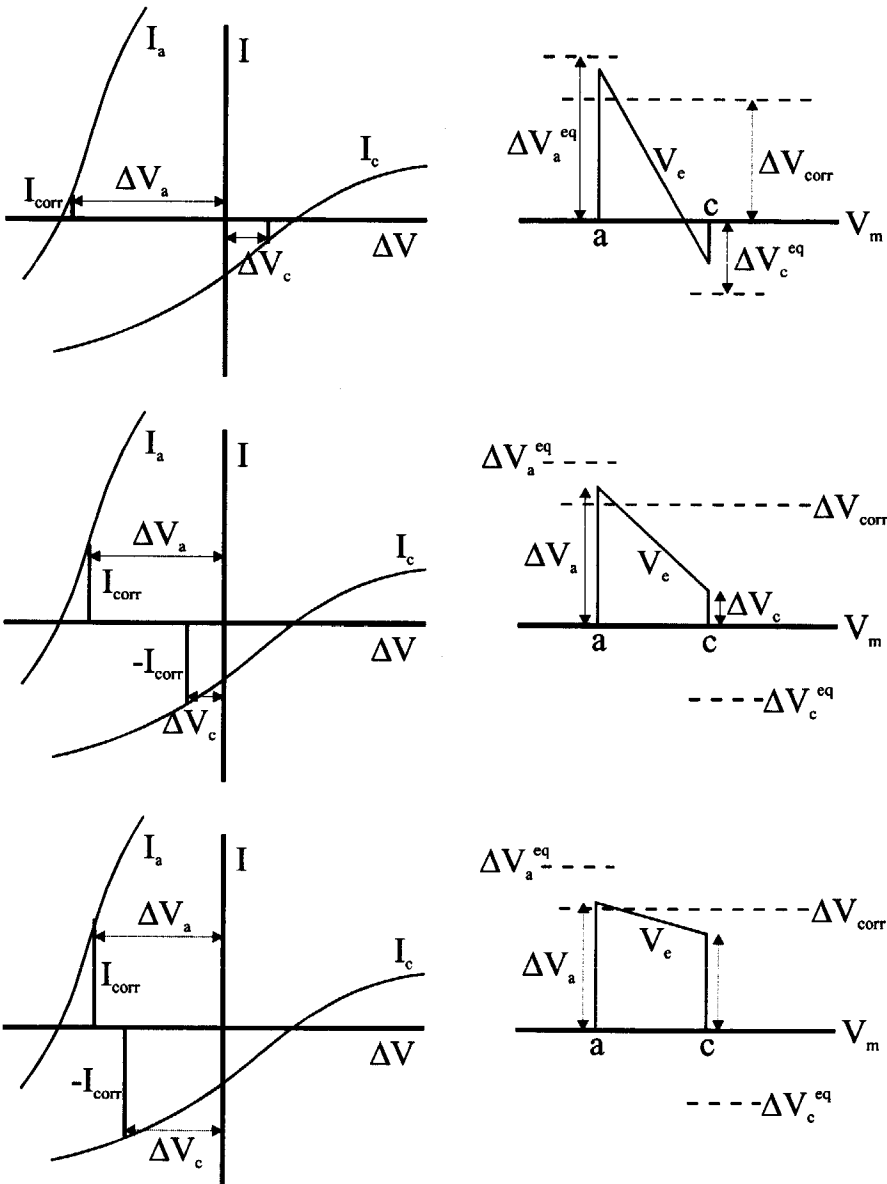
In case a considerable corrosion current I_{corr} runs from the anodic to the cathodic site, the anodic and cathodic sites are no longer in equilibrium. The magnitude of I_{corr} is then determined by the difference between ΔV_a^{eq} and ΔV_c^{eq} (similar to uniform corrosion) and by the conductive coupling between the anodic and cathodic sites. The latter is illustrated by the figures 5.5a to 5.5f, which are three pairs of figures in which the conductivity increases from the higher to the lower figures. The left figures are similar to figure 5.3 and show the magnitude of the corrosion current as well as the departure from the equilibrium voltages ΔV_a^{eq} and ΔV_c^{eq} . The left figures also show the potential drop in the electrolyte between the anodic and cathodic sites, which is the difference between ΔV_a and ΔV_c . In the right figures, the electric potential in the electrolyte V_e is plotted as a function of the distance between the anodic and cathodic sites. It follows that for a higher conductive coupling between the sites:

- The corrosion current I_{corr} is higher;
- The electric potential V_e is more flat.

Eventually, when the conductivity between the sites becomes very high, the potential differences in the electrolyte become negligible and $V_m - V_e$ is more or less equal to the corrosion potential ΔV_{corr} , for the whole electrolyte between the anodic and cathodic sites.

We see that in the first place, potential scans can be used for detecting anodic and cathodic sites on a corroding surface, and this alone may already provide a lot of valuable information. However, it also follows that the magnitude of the potential differences is not related to the localized corrosion rates. In fact, a flatter potential scan can mean that the local corrosion rates are high, but could of course also correspond to more uniform corrosion processes, or even to the absence of corrosion activity. It is therefore difficult to deduce more from potential scans than the detection of anodic and cathodic sites, which is indeed already interesting by itself.

For the interpretation of potential scans, one must also consider other phenomena influencing the electric potential on a surface. Below, and particularly in section 5.2.4, we will discuss a number of these phenomena.

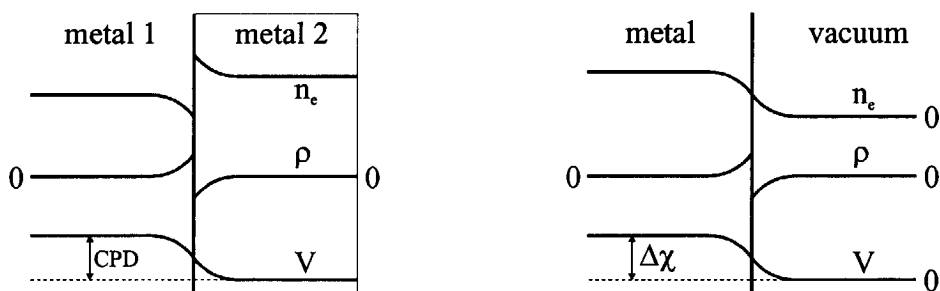


Figures 5.5a-f. The corrosion current and electric potential in the electrolyte as a function of the conductivity of the electrolyte. Going from top to bottom, the conductivity of the electrolyte increases. The left figures show the corrosion currents and the right figures show the electric potential when going from the anodic site (a) to the cathodic site (c).

5.2.3 The electric potential in the bulk around interfaces

In the previous sections, it was assumed that the surface properties of metals and electrolytes do not differ from the bulk properties. In this picture, the chemical potential of a homogeneous material is constant from the bulk to the surface and the electric potential really makes a discrete jump at the interface between a metal and an electrolyte or at the interface between two different metals. This was sufficient for the discussion on the influence of corrosion on the electric potential in the above sections, but for discussing the influence of number of other effects, the “bulk treatment” must be somewhat refined. (We will do this only as much as necessary. In fact, one can always proceed to smaller scales.)

First we reconsider the situation at the interface between two metals with different μ_e^{int} as in figure 5.2. In section 5.2.1 it was stated that in this case, the electric potential makes a discrete jump at the interface with magnitude equal to the CPD. However, this would imply that the charge density at the interface would jump from $+\infty$ to $-\infty$, which is physically impossible. In reality, the internal chemical potential can no longer be regarded as independent of the electron density at such extreme electron densities. This leads to a finite “double layer” of positive and negative charges around the interface. By consequence, the electric potential does not make a really discrete jump and the CPD will only be reached between points relatively deep in the bulk of the metals, as illustrated by figure 5.6a. For the interface between a metal and an electrolyte, similar considerations hold. There will also be a finite double layer, and the corrosion voltage over the metal-electrolyte interface will only be reached between two points in the bulk of the metal and the electrolyte.



Figures 5.6a and 5.6b. The electron density n_e , the charge density ρ and electric potential V around the interface between two metals and at a metal surface

Also the electric potential around the interface between a metal and vacuum is of importance and needs a closer look. In this case, one must consider the effect that the electron wave function “leaks out” into the vacuum. By consequence, the vacuum is somewhat negatively charged, while the metal is somewhat positively charged, as shown in figure 5.6b. This resulting double layer is called the “surface dipole” and leads to a decrease of the electric potential when going from the metal to the vacuum (cf. [Zangwill 1988]). The voltage associated with the surface dipole, which is denoted by $\Delta\chi$, depends on the local crystal structure at the surface. This can lead to considerable differences. For gold for instance, $\Delta\chi$ differs by 160mV between the (100) and (111) surfaces [Weast 1975].

Not only for metals, but also for other materials, surface dipoles play a role because the electronic structure at the surface of a material is usually different from the bulk. The surface dipole of a diluted electrolyte is generally regarded as independent of the dissolved ions and molecules (cf. [Stratmann 1987]). The idea behind this assumption is that the surface of a diluted electrolyte mainly consists of solvent molecules. However, this may be different when the solvent contains surface-active agents, which preferably accumulate on the surface. This will affect $\Delta\chi$.

5.2.4 Coatings

In sections 5.2.1 and 5.2.2, metal substrates with or without electrolyte layers were considered, with special attention for the relation between the electric potential and corrosion activity. However, the electric potential can be influenced by other phenomena than corrosion, especially for metal substrates with coatings. For metallic or semi-conducting coatings, the discussion of section 5.2.1 suffices to describe the spatial dependence of the electric potential above the substrate, but the case of insulating or poorly conducting coatings is more complicated. The following effects then have to be taken into account:

- The surface dipole of the coating;
- Static charges inside or on top of the coating;
- The potential difference over the substrate-coating interface;
- The small conductivity of the coating.

The effect of the surface dipole of the coating is covered by the discussion in section 5.2.3. The effect of static charges is that, in general, positive static charges raise and negative static charges lower the electric potential above a substrate. The other two effects need some more discussion.

In case the coating has no electronic conductivity, there will be no charged particle equilibrium between the coating and the metal and there will be no CPD. Nevertheless one may expect a potential difference over the metal-coating interface in case of a chemical binding between the metal and the coating. This will often be the case in practice, since the adhesion of a coating generally relies on such a binding. The effect on the electric potential can be compared to the effect of a layer of absorbed molecules on a metal. This latter case is more extensively studied by means of work function measurements (cf. [Zangwill 1988]). For instance, the adsorption of one monolayer of xenon atoms changes the work function of copper by 0.6eV [Chen 1984], corresponding to a change of 0.6V of $\Delta\chi$ and the adsorption of a monolayer of chlorine atoms decreases $\Delta\chi$ by even 1.4V [Goddard 1977]. This stronger effect can be understood from the higher electronegativity of chlorine.

The effect of the coating conductivity and of the electric and dielectric behavior of the coating in general is very difficult to estimate. One effect of a small coating conductivity is that possible potential differences underneath the coating can be leveled out and be hidden for potential scanners. Whether this happens and to what extent it happens is very hard to predict, since the electric and dielectric properties of coatings are often not so easily described.

5.3 Summary and conclusions

This study was motivated by the recent introduction of a number of new techniques in corrosion research, such as the SRET (scanning reference electrode technique), the SVET (scanning vibrating electrode technique) and the SKP (scanning Kelvin probe). These techniques have in common that they scan the electric potential above a corroding surface. An important practical difference between the techniques is that with the SRET and the SVET the electric potential is scanned in the electrolyte, while the SKP scans the electric potential just above the electrolyte. Therefore the SKP is also used for studying atmospheric and underpaint corrosion.

The electric potential above a corroding surface gives information on the corrosion processes occurring, namely the location of anodic and cathodic sites. In this chapter we have studied whether it is also possible to deduce the local corrosion rates from the measured potential differences. In a way, one might expect this to be possible, because a metal-electrolyte system is no longer in thermodynamic equilibrium when a corrosion current flows and the deviation from the equilibrium potential depends on the corrosion current. However, already for the simplest corroding systems, obeying the Stern-Geary approximation, the corrosion current and the corrosion potential depend on too many parameters to result in a one on one relation between them.

We also studied the influence of the conductive coupling between anodic and cathodic sites on the potential above a corroding surface. We found that for a higher conductive coupling, the corrosion current is higher, while the potential difference between the anodic and cathodic sites is smaller. Eventually for a perfect coupling between the sites, when the localized corrosion rates are maximal, the potential in the whole electrolyte becomes equal to the corrosion potential, and the potential scan gives the same result as in case of uniform corrosion. This contradiction illustrates the difficulty of interpreting potential scans for more than only the location of anodic and cathodic sites.

In addition, the electric potential above a corroding surface is also influenced by other phenomena than corrosion:

- The metal composition of the surface;
- Surface dipoles of metals, electrolytes and coatings;
- The chemical adhesion of adsorbed layers or coatings;
- Static charges on insulating coatings.

It follows that especially for studying the corrosion of coated metals there can be many disturbing effects. (Here, we think of insulating coatings such as oxide layers or polymer coatings.) In addition, the voltages associated with the mentioned effects can be in the order of volts or more, while the potential differences associated with corrosion phenomena are typically in the order of 100mV. This means that experiments with potential scanners have to be done with great care and patience.

The above list of phenomena influencing the surface potential also suggests several other applications of potential scanners, especially for the SKP. For instance one could think of using the SKP for measuring the influence of surface-active agents in electrolytes on the surface dipole, or one could use the SKP for studying the adhesion strength of coatings on

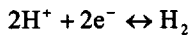
metals. The latter application is relevant to corrosion research and could imply a new method for studying local adhesion differences of coatings. Finally, the SKP can also be used for charge decay measurements of coatings, which also provides a way of studying coatings on metals. The versatility of the SKP was the reason that for this work, we especially focus on this potential scanner. The results of our investigations are discussed in the next chapter.

Appendix 5A. The Nernst equation

In this appendix we will give a compact but complete derivation of the Nernst equation, following the methods and notations of Kittel [Kittel 1980]. The Nernst equation is an expression for the voltage over a metal-electrolyte interface where some chemical reaction is in thermodynamic equilibrium. Suppose that one can write this chemical reaction in the following general form:

$$\sum_j \nu_j A_j \leftrightarrow 0$$

where A_j denotes the reaction species with index j and ν_j is the coefficient of the species in the reaction. For instance, for the following chemical reaction:



one can choose $\nu_1 = \nu(\text{H}^+) = 2$, $\nu_2 = \nu(\text{e}^-) = 2$ and $\nu_3 = \nu(\text{H}_2) = -1$ as the reaction coefficients. (Equivalently, one could have chosen the coefficients -2 , -2 , 1 or 4 , 4 , -2 and so on. This leads to the same final result) At constant temperature and pressure, the Gibbs free energy G is conserved, hence:

$$dG = \sum_j \mu_j dN_j = 0$$

Here, μ_j is the chemical potential of species j and N_j is the number of molecules of each species. Now the above reaction equation implies that the changes dN_j are proportional to the reaction coefficients of the species ν_j . (In fact, when the reaction takes place dN times, $dN_j = \nu_j dN$ for all j .) This leads to the following:

$$\sum_j \nu_j \mu_j = 0$$

In the absence of electric (and other) potential differences, the electrochemical potentials are equal to the internal chemical potentials:

$$\mu_j = \mu_j^{\text{ref}} + k_B T \ln(n_j / n_j^{\text{ref}})$$

where n_j is the density of species j . For ions in electrolytes, the usual reference concentrations are 1 mole per liter, while for gases in electrolytes, one usually takes the saturation concentration at 1 atm. When the solvent of the electrolyte takes part in the reaction, its chemical potential must be considered too. Then a choice for the reference concentration must be made. The choice is usually the concentration of the pure solvent. The concentration dependent term of the solvent can then be discarded for dilute solutions. Substitution of the internal chemical potentials in the above equilibrium condition gives:

$$\sum_j \nu_j \mu_j^{\text{ref}} + k_B T \sum_j \nu_j \ln(n_j / n_j^{\text{ref}}) = 0$$

This can be rewritten in the more familiar form:

$$\prod_j \left(\frac{n_j}{n_j^{\text{ref}}} \right)^{\nu_j} = \exp \left[- \frac{\sum_j \nu_j \mu_j^{\text{ref}}}{k_B T} \right] \equiv K_{\text{eq}}(T)$$

where K_{eq} is the equilibrium constant of the chemical reaction.

When the electrolyte is in contact with a metal electrode and when electrons from the electrode take part in the reaction, the electric potentials V_m and V_e of the electrode and the electrolyte must also be considered. Then the electrochemical potentials must be used:

$$\mu_j = \mu_j^{\text{ref}} + k_B T \ln(n_j / n_j^{\text{ref}}) + q_j V_j$$

Here, V_j is the electric potential that species j experiences. When the metal atoms in the electrode take part in the reaction, their chemical potential must be considered too. However, their charge is zero and since the concentration of metal atoms in the electrode is fixed, their chemical potential is always the same. For the chemical potential of the electrons in the metal electrode one can write:

$$\mu_e = \mu_e^{\text{ref}} + q_e V_m$$

All other particles taking part in the chemical reaction will be found in the electrolyte, hence one can write the following for their chemical potentials:

$$\mu_j = \mu_j^{\text{ref}} + k_B T \ln(n_j / n_j^{\text{ref}}) + q_j V_e$$

Here the concentration dependence is significant, because the ion concentrations can be influenced by dissolving salts while the gas concentrations can be influenced by changing the partial pressure of the particular gas specimen above the electrolyte. Note that the charge q_j can be zero for some of the species. Substitution of all chemical potentials into the equilibrium condition gives:

$$v_e \mu_e^{\text{ref}} + v_e q_e V_m + \sum_{j \neq e} v_j [\mu_j^{\text{ref}} + k_B T \ln(n_j / n_j^{\text{ref}})] + \sum_{j \neq e} v_j q_j V_e = 0$$

This is in fact as good as the final result. However, the equation can be written in a more convenient form containing the voltage between the metal and the electrolyte instead of the actual potential values. For that, we note that the conservation of electric charge during the reaction implies:

$$v_e q_e + \sum_{j \neq e} v_j q_j = 0$$

Hence one can write the following:

$$\sum_{j \neq e} v_j q_j V_e = -v_e q_e V_e$$

Substitution in the previously found result gives:

$$v_e q_e (V_m - V_e) + v_e \mu_e^{\text{ref}} + \sum_{j \neq e} v_j [\mu_j^{\text{ref}} + k_B T \ln(n_j / n_j^{\text{ref}})] = 0$$

After some minor manipulations this can be rewritten as:

$$V_m - V_e = (V_m - V_e)^{\text{ref}} + \frac{k_B T}{v_e |q_e|} \ln \left[\prod_{j \neq e} \left(\frac{n_j}{n_j^{\text{ref}}} \right)^{v_j} \right]$$

This equation is called the Nernst equation. It gives the voltage between the metal and the electrolyte as a function of the relevant particle concentrations. When all concentrations would be equal to the reference concentrations, the voltage is equal to the reference voltage $(V_m - V_e)^{\text{ref}}$. Tables of reference voltage for different chemical reactions can be found in many handbooks (see for example [Weast 1975]). In fact, Nernst originally only considered the equilibrium of one metal ion at the electrode [Nernst 1889]. The extension to several particles originates from Gibbs (cf. [Dubpernell 1978]).

Appendix 5B. The Butler-Volmer equation

In this appendix we give a semi-classical derivation of the Butler-Volmer equation, which is a relation for the current of electric charge carriers through a potential barrier as a function of the voltage across the barrier.

In the usual semi-classical treatment of particle flows through potential barriers (cf. [Feynman 1963]) it is assumed that the fraction of particles crossing the potential barrier is proportional to the fraction of particles that have sufficient kinetic energy for doing so. For non-interacting particles, this fraction follows from the Boltzmann distribution of the kinetic energy E :

$$p(E) \propto \exp[-E/k_B T]$$

Now the fraction of particles that are able to cross a given potential barrier A reads:

$$p(E > A) = \frac{\int_{E=A}^{\infty} \exp[-E/k_B T] dE}{\int_{E=0}^{\infty} \exp[-E/k_B T] dE} = \exp[-A/k_B T]$$

This is in fact a one-dimensional treatment of the kinetic energy, justified by the fact that we consider a one-dimensional barrier such as a metal-electrolyte interface. (Hence, one only needs to consider the distribution of one component of the particle velocity, which is the component perpendicular to the interface.) When the height of the potential barrier is A_1 for particles crossing from side 1 and A_2 from side 2, one can write the following for the net current through the potential barrier:

$$I = qn_1 k_{12} \exp[-A_1/k_B T] - qn_2 k_{21} \exp[-A_2/k_B T]$$

This is the net difference between the currents from either side. The reaction rates k_{12} and k_{21} are the fractions of the particles per unit of time that not only have sufficient kinetic energy to cross the barrier, but that also actually do so. This expression is in fact equivalent to the Butler-Volmer equation. The rest of the discussion is mostly a matter of rewriting this result in the most convenient form, that is to write I as a function of the voltage across the barrier.

First we consider the equilibrium state of the system in which the net current I through the barrier is zero. This leads to the following expression:

$$\left(\frac{n_2}{n_1}\right)_{\text{eq}} = \frac{k_{12}}{k_{21}} \exp\left[\frac{(A_2 - A_1)_{\text{eq}}}{k_B T}\right]$$

Now the equilibrium can also be described by considering the electrochemical potentials on either side of the barrier:

$$\begin{aligned} \mu_1 &= \mu_1^{\text{ref}} + k_B T \ln(n_1/n_1^{\text{ref}}) + qV_1 \\ \mu_2 &= \mu_2^{\text{ref}} + k_B T \ln(n_2/n_2^{\text{ref}}) + qV_2 \end{aligned}$$

where q is the charge per particle and V_1 and V_2 are the electric potentials. The reference concentrations refer to the electrically neutral states. In equilibrium, it holds that $\mu_1 = \mu_2$, from which the following can be derived:

$$\left(\frac{n_2}{n_1}\right)_{\text{eq}} = \left(\frac{n_2^{\text{ref}}}{n_1^{\text{ref}}}\right) \exp\left[-\frac{(\mu_2^{\text{ref}} - \mu_1^{\text{ref}}) + q(V_2 - V_1)_{\text{eq}}}{k_B T}\right]$$

and we have again an expression for $(n_2/n_1)_{eq}$. Now it is interesting to compare both expressions. For that purpose, we must write the activation energies in terms of the electrochemical potentials and the voltages. Recall that the activation energy is the kinetic energy required for crossing the potential barrier. Hence, when the height of the potential energy barrier is W_b , and the potential energies in 1 and 2 are W_1 and W_2 , the activation energies read:

$$\begin{aligned} A_1 &= W_b - W_1 \\ A_2 &= W_b - W_2 \end{aligned}$$

Now the difference between the chemical potential and the potential energy is just the concentration dependent term:

$$W = \mu_{ref} + qV$$

(Note that the reference chemical potential is also a potential energy contribution, referring to the electrically neutral state $V = 0$ by the above definition.) With that, one can write the following for the activation energies:

$$\begin{aligned} A_1 &= W_b - (\mu_1^{ref} + qV_1) \\ A_2 &= W_b - (\mu_2^{ref} + qV_2) \end{aligned}$$

Using these expressions for the activation energies, one can now compare the two above results for the ratio $(n_2/n_1)_{eq}$. This leads to the following nontrivial result for the reaction rate constants:

$$\frac{k_{12}}{k_{21}} = \frac{n_2^{ref}}{n_1^{ref}}$$

Also one can derive the following condition for the equilibrium potentials:

$$qn_1^{eq}k_{12} \exp\left[-\frac{W_b - (\mu_1^{ref} + qV_1^{eq})}{k_B T}\right] = qn_2^{eq}k_{21} \exp\left[-\frac{W_b - (\mu_2^{ref} + qV_2^{eq})}{k_B T}\right] \equiv I_0$$

By this equation, the equilibrium voltage ($V_1^{eq} - V_2^{eq}$) over the barrier is defined. In fact, this relation is equivalent to the Nernst equation.

We will now turn back our attention to the net current I through the interface in the non-equilibrium situation. Using the above expressions for the activation energies in terms of the electric potential, one can derive the following:

$$I = I_0 [\exp(q\Delta V_1 / k_B T) - \exp(q\Delta V_2 / k_B T)]$$

Here, ΔV_1 and ΔV_2 are the departures of the equilibrium potentials on either side of the potential barrier. That is to say:

$$\begin{aligned}V_1 &= V_1^{\text{eq}} + \Delta V_1 \\V_2 &= V_2^{\text{eq}} + \Delta V_2\end{aligned}$$

The actual voltage over the potential barrier reads:

$$V_1 - V_2 = (V_1^{\text{eq}} - V_2^{\text{eq}}) + (\Delta V_1 - \Delta V_2)$$

The first part of the right hand side of the equation is just the equilibrium or Nernst voltage, leading to a zero net current through the potential barrier. The second part is therefore the voltage that would be applied by an external voltage source:

$$\Delta V_{\text{ext}} = \Delta V_1 - \Delta V_2$$

It is this voltage that leads to a nonzero net current I through the interface and we will see that it is possible to write I as a function of ΔV_{ext} .

The external voltage is shared over both sides. For instance, if $\Delta V_{\text{ext}} = 1\text{V}$, it could be that the electric potential on side 1 is raised by 0.5V while the electric potential on side 2 is lowered by the same amount. Then $\Delta V_1 = 0.5\text{V}$ and $\Delta V_2 = -0.5\text{V}$ and the voltage is equally shared over both sides. In that case, one speaks about a symmetric potential barrier. But strictly, this needs not to be the case. In fact, the external voltage could also be shared by:

$$\begin{aligned}\Delta V_1 &= \alpha \Delta V_{\text{ext}} \\ \Delta V_2 &= -(1 - \alpha) \Delta V_{\text{ext}}\end{aligned}$$

where the value of the parameter α lies between 0 and 1. Substitution into the expression for the net current through the barrier gives:

$$I = I_0 \left[\exp[\alpha q \Delta V_{\text{ext}} / k_B T] - \exp[-(1 - \alpha) q \Delta V_{\text{ext}} / k_B T] \right]$$

This equation is known as the Butler-Volmer equation. It appears first in an article by Erday-Grúz and Volmer [Erday-Grúz 1930], but before these authors, Butler had also derived exponential expressions for the current through a metal-electrolyte interface [Butler 1924]. In both articles, the derivation is based on the empirical result of Tafel that the logarithm of the modulus of the current through a metal-electrolyte interface as a function of ΔV_{ext} gives asymptotically a straight line [Tafel 1900]. Since 1930, various authors have improved the semi-classical derivation (cf. [Bockris 1970]). It is interesting that the Butler-Volmer equation also occurs (on different names) in many other research fields, for instance in the theory of transistors and in the theory of the ion conductivity of insulators by particle hopping. The reason is that in all these cases charged particles are crossing potential barriers.



Chapter 6

Measuring with the Scanning Kelvin Probe

6.1 Introduction

The Kelvin probe (KP) measures the electric potential on a surface with respect to a metal probe or reference electrode. When equipped with a probe with a fine tip, the apparatus can be used for measuring lateral differences of the electric potential on surfaces. Then the apparatus is referred to as the scanning Kelvin probe (SKP). As discussed in the previous chapter, the electric potential on a surface is influenced by many factors and therefore the KP and SKP are used in many research fields. For instance, the apparatus is used for studying:

- The metal composition of surfaces [Seminikhin 1997, Schmutz 1998];
- Doping distributions in semiconductors [Kikukawa 1996, Sakalauskas 1998];
- Wear precursors on metals [DeVecchio 1998, Kasai 1998];
- The influence of adsorbed layers on the work function of metals [Baikie 1998];
- Stress on metals [Craig 1969];
- Charge decay as a function of temperature on polymers [Schaeffert 1970];
- The chemical potential of high T_c superconductors [Rietveld 1993];
- Atmospheric corrosion [Tahara 1997, Wang 1997, Chen 1998, Nazarov 1998];
- Underpaint corrosion [Stratmann 1991].

In the previous chapter the SKP was estimated as the most interesting electric potential scanner for our purposes, and therefore in this chapter the SKP is studied in more detail. In section 6.2, the Kelvin probe method of measuring electric potentials on surfaces is discussed. The resolution of the SKP will be discussed as well as several methods of controlling the distance between the probe and the substrate during a scan. Special attention will be paid to the influence of insulating or poorly conducting coatings. Finally in section 6.3, a number of SKP measurements on typical systems is presented to verify the ideas and predictions of section 6.2 and the previous chapter.

6.2 The Kelvin probe method

One may perhaps wonder why potential differences on metal surfaces are not simply measured by connecting the clamps of a voltmeter to different positions on the surface. The reason is that ordinary voltmeters measure a voltage by means of an electric current caused by the voltage. But when the clamps of a voltmeter are connected to different points on a metal surface, no steady current will flow because the sum of the CPDs (contact potential differences) in the circuit will be zero. As a result, the voltmeter will always indicate zero.

It follows that a special tool is needed to measure potential differences on surfaces and the Kelvin probe is such a tool. The apparatus is based on the ideas of Lord Kelvin who demonstrated the existence of CPDs by a series of experiments and papers [Kelvin 1898]. In the first subsection, we will discuss the non-scanning version and then in a number of subsequent subsections, the scanning apparatus as well as a number of related topics.

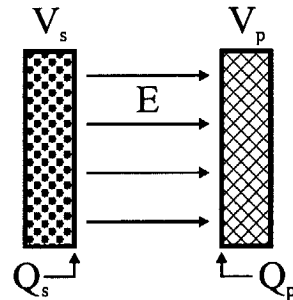
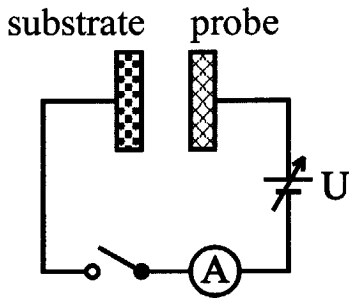
6.2.1 Homogeneous surfaces

Lord Kelvin showed the existence of the CPD by connecting a large zinc plate to a copper plate. After the contact was made, electrons flow to the copper plate and both plates are charged. The charges remained when the plates were disconnected and Kelvin showed the presence of the charges by connecting the plates to the housing and the gold leaves of an electroscope. Zisman, following a suggestion of Prof. Knoble, considerably improved the above method of measuring CPDs [Zisman 1932]. His experimental setup is sketched in figure 6.1. With this setup, the CPD between two metals, the substrate and the probe is measured. The circuit contains a direct voltage source U and an amperemeter A . We will now discuss the measurement in detail.

First we consider the situation when $U = 0$ and with the switch open. For simplicity it will be assumed that the left part of the circuit, from the switch to the substrate, is entirely made of the substrate metal and that the right part of the circuit is entirely made of the probe metal. Then as long as the switch is open, all objects are electrically neutral, the electric potential is zero everywhere, and the electric field between the plates is zero. Now let us consider the situation where the compensation voltage U is still zero, but with the switch closed. Following the discussion of section 5.2.1 of the previous chapter, this leads to a CPD over the switch and hence to a potential difference between the probe and the substrate equal to the CPD. This means that there is an electric field between the probe and the substrate plates and in case the plates are large, flat and parallel, this electric field reads:

$$E = -\frac{dV}{dz} = -\frac{V_p - V_s}{z_p - z_s} = -\frac{V_{ps}}{z_{ps}}$$

where V_{ps} is equal to the CPD and z_{ps} is the mutual distance between the probe and the substrate. In the following, the z -coordinate is defined as the distance from the substrate and increases in the direction of the probe. The electric field E is positive when directed from the substrate to the probe, as shown in figure 6.2.



Figures 6.1 and 6.2. Kelvin probe experimental setup and the electric field between the substrate and the probe.

Inside the substrate and the probe, the electric field is zero, and at the positions where the electric field E originates (the substrate) and disappears (the probe), there must be electric charges. These charges, which will be denoted by Q_s and Q_p , are proportional to E , as described by Gauss's law:

$$Q_s = \epsilon A E = \frac{\epsilon A}{z_{ps}} V_{ps} \equiv C_{ps} V_{ps} \tag{6.1a}$$

$$Q_p = -\epsilon A E = -C_{ps} V_{ps} \tag{6.1b}$$

Here, ϵ is the permittivity of the medium between the substrate and the probe, and A is the area of the facing surfaces of the substrate and the probe. Since the electric field originates in the substrate and disappears in the probe, Q_s has the same sign as E and Q_p has opposite sign. In equation (6.1a), the capacity C_{ps} between the plates is defined as $\epsilon A/z_{ps}$. This is the well-known expression for the double plate capacitor.

Now by the voltage U , the CPD can be compensated such that $V_{ps} = 0$, $E = 0$ and $Q_s = Q_p = 0$. When this is the case, U is equal to the CPD, hence the CPD is measured by nulling the charges on the plates. In this way, Kelvin measured CPDs. But instead of measuring the charges, Zisman continuously modulated the mutual distance z_{ps} between the plates. Thus the mutual capacity C_{ps} between the plates is modulated, and by consequence the charges on the plates. Through the circuit then runs the following alternating current, transporting charge from and to the plates:

$$I(t) = \frac{dQ_s}{dt} = V_{ps} \frac{dC_{ps}}{dt} = (CPD - U) \frac{dC_{ps}}{dt} \tag{6.2}$$

Zisman measured the amplitude of $I(t)$ by amplification and using headphones. This came down to adjusting U such that the sound volume was minimized and could be done within seconds and with an accuracy of 1mV. Nowadays, a lock in amplifier (LIA) is more commonly used to measure the amplitude of $I(t)$.

Since the electric field E between the substrate and the probe is proportional to the charges Q_s and Q_p , one can equivalently say that the CPD is measured by nulling the electric field between the substrate and the probe. This makes no difference in “bulk treatment” given above, but leads to a more convenient picture on a microscopic scale, where it is not so clear what and where the charges Q_s and Q_p are, nor what electric potential is felt by them. For instance it is then difficult to say whether the potential difference $\Delta\chi$ due to the surface dipole is also measured or not. By considering the electric field E instead, these details are avoided and it is immediately clear that the surface dipole contributions for both the substrate and the probe are included in the measured CPD (which is the U value that nulls the amplitude of the alternating current). Therefore the measured CPD is not exactly the same as the potential difference between the bulk of the substrate and the probe.

In the above, the “backward contact” between the substrate and the probe was made by closing a switch. In reality, the backward contact will be made when the substrate is mounted in the setup. This makes of course no difference for the discussion. In principle, it even does not matter whether the backward circuit contains other metals since in thermodynamic equilibrium, the CPDs in the circuit will add up to the CPD between the probe and the substrate materials. In practice however, it may be questionable whether the setup is really in thermodynamic equilibrium. Especially when thin wires are used or when the electric contact to the substrate is not very good, the heat produced by the electronics may not be carried away properly and the circuit may not be at one and the same temperature. A badly controlled temperature may lead to slow drifting of the measured U and to irreproducible measurements.

We will now work out a more detailed expression for the probe current $I(t)$ for later use, assuming that the probe is vibrated harmonically. The probe is usually vibrated electrically, by mounting it to an electric actuator such as a loudspeaker or a piezo. When ω is the vibration frequency and δz is the vibration amplitude, the height of the probe above the substrate as a function of time reads:

$$z(t) = z_0 + \delta z \sin(\omega t)$$

Here $z_0 = z_{ps}$ is the distance between the probe and the substrate in rest. Substitution in expression (6.2) for the probe current gives:

$$I(t) = V_{ps} \frac{dC_{ps}}{dt} = V_{ps} \frac{dC_{ps}}{dz} \frac{dz}{dt} = V_{ps} C_{ps}'[z(t)] \omega \delta z \cos(\omega t)$$

For the double plate geometry, a closed expression for $C_{ps}'(z)$ can be given, leading to an analytical expression for $I(t)$. Instead we will give first and second order approximations, which are also valid for other probe-substrate geometries and more suited to describe the actual measurement of $I(t)$ by a LIA. For that purpose, the following Taylor-expansion for $C_{ps}'[z(t)]$ will be used:

$$C_{ps}'[z(t)] = C_{ps}'[z_0 + \delta z \sin(\omega t)] \cong C_{ps}'(z_0) + C_{ps}''(z_0) \delta z \sin(\omega t)$$

Substitution in the above expression for $I(t)$ gives:

$$I(t) \cong V_{ps} C_{ps}'(z_0) \omega \delta z \cos(\omega t) + V_{ps} C_{ps}''(z_0) \delta z \sin(\omega t) \omega \delta z \cos(\omega t)$$

By some elementary trigonometry, this can be rewritten in the following form:

$$I(t) \cong V_{ps} \omega \delta z \left[C_{ps}'(z_0) \cos(\omega t) + \frac{\delta z}{2} C_{ps}''(z_0) \sin(2\omega t) \right] \quad (6.3)$$

Hence in the lowest order, the current response only contains a $\cos(\omega t)$ term and is therefore linear. The first order distortion of the probe current is a second harmonic term. Higher order approximations of $I(t)$ can be obtained and will also contain harmonics of ω , since $I(t)$ must have the same period as the probe vibration. Now when the amplitude of $I(t)$ is measured by a LIA, only the amplitude of the $\cos(\omega t)$ term is measured and the lowest order approximation exactly describes the measurement. The second harmonic can also be measured by a LIA and in section 6.2.3, we will see an application of this fact.

6.2.2 Inhomogeneous surfaces

For inhomogeneous metal surfaces on which the electric potential laterally varies, there is a different CPD between the probe and different points of the surface and one can say that the CPD laterally varies. In this section we will discuss what the SKP measures in this case, hence for what voltage U the probe current is nulled. From intuition one might guess the following:

- One measures the average of the CPDs under the probe;
- The CPDs closest to the probe are dominant.

We will see that this is indeed more or less the case. In the following discussion, we will work out these statements quantitatively.

First we consider the case where the lateral variations of the surface potential are modest with respect to the dimension of the probe tip. We will see that under some well-defined conditions, the major part of the discussion of section 6.2.1 is then still valid. The main difference is that the expression for C_{ps} , defined in (6.1a) is not valid for a probe with a fine tip. However, in case the major part of the electric field lines flows between the probe and the substrate or in other words, when there is no leakage of electric field lines to the environment, one can still write an expression of the following form (cf. [Smythe 1968]):

$$Q_s = C_{ps} V_{ps} = -Q_p$$

where C_{ps} is the mutual capacity between the probe and the substrate. Hence under the given constraints, the charges on the probe and the substrate are still opposite and proportional to the potential difference V_{ps} between the probe and the substrate. The mutual capacity C_{ps} is determined by the geometry and still depends on the mutual distance z_{ps} between the probe and the substrate. Therefore one can still write the following for the probe current:

$$I(t) = \frac{dQ_s}{dt} = (CPD - U) \frac{dC_{ps}}{dt}$$

Also the second order approximation of $I(t)$, given by (6.3), still applies. Hence under the given constraints (no leakage of the electric field), the probe current is still nulled by $U = CPD$ and the SKP measures the local CPD.

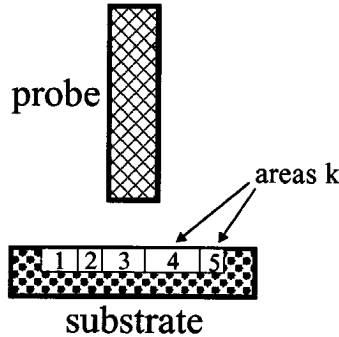


Figure 6.3. The SKP above an inhomogeneous surface. The surface of is divided in a number of different areas with different electric potentials

Now suppose that the CPD on the substrate considerably varies compared to the dimensions of the probe tip. This case can be considered by dividing the surface of the substrate into a large number of small areas as shown in figure 6.3. The areas do not need to have the same form or size. Now in case the potential differences between the probe and area k reads V_{pk} , the charge on the probe reads:

$$Q_p = -\sum_k C_{pk} V_{pk} \quad (6.4)$$

Here C_{pk} is the mutual capacity between the probe and area k . The sum of all C_{pk} is equal to the capacity C_{ps} between the probe and the substrate, as follows from considering a homogeneous substrate with V_{pk} equal to V_{ps} for all k . Equivalently, the latter equation could be written as a two-dimensional surface integral, but the discrete expression is also complete and simpler. Now when the probe is vibrated, all C_{pk} are modulated and the following probe current will flow through the circuit:

$$I(t) = \left[\sum_k C_{pk}'(z) V_{pk} \right] \frac{dz}{dt} \cong \left[\sum_k C_{pk}'(z_0) V_{pk} \right] \frac{dz}{dt}$$

For a measurement, this current is nulled, giving:

$$\sum_k C_{pk}'(z_0)V_{pk} = 0$$

Now when CPD_k is the CPD between the probe and area k on the substrate, we have $V_{pk} = CPD_k - U$, and it follows that $I(t)$ is nulled by:

$$U = \frac{1}{\sum_k C_{pk}'(z_0)} \sum_k C_{pk}'(z_0)CPD_k = \frac{1}{C_{ps}'(z_0)} \sum_k C_{pk}'(z_0)CPD_k \quad (6.5)$$

Here, C_{ps} is the capacity between the probe and the substrate when it would be homogeneous. It follows that U is equal to a weighed average of the CPD values on the surface, where the weight factor of area k is equal to $C_{pk}'(z_0)/C_{ps}'(z_0)$. For instance for a non-scanning Kelvin probe, for which the probe is a plate parallel to the substrate, all $C_{pk}'(z_0)$ are equal and U is simply equal to the average CPD on the surface. For a SKP, the $C_{pk}'(z_0)$ for the area right under the probe tip will be the largest and hence the area right under the probe tip has the largest contribution to the measured compensation voltage U .

Thus, the intuitive statements of the beginning of this section are quantified. This section is completed by two discussions on related topics. The first discussion is on the enhancement of potential scans by inversion techniques and the second discussion is about the influence of electric field leakage to the environment.

6.2.2.1 Data enhancement by inversion techniques

The weight factors $C_{pk}'(z_0)/C_{ps}'(z_0)$ provide detailed information on the apparatus and once they are known, they can be used for a better interpretation of the measured potentials. In the following, the factors $C_{pk}'(z_0)/C_{ps}'(z_0)$ will be called the "response function" and denoted by h_k . With that, equation (6.5) can be rewritten as:

$$U = \sum_k h_k CPD_k$$

The response function contains about all relevant information of the apparatus. For instance, the width of the response function is the lateral resolution. For an ideal apparatus, $h_k = 1$ for the area right under the probe and zero elsewhere. Then U is equal to the CPD right under the probe. In practice however, h_k will have a finite width, which is a measure for the lateral resolution. Analytical expressions for h_k will never be available in practice, but h_k can also be obtained from numerical calculations or from a measurement. It may payoff to spend some effort in obtaining h_k , because it can be used to enhance the potential scan by inversion techniques. We will not go into too much detail here, but only mention the main ideas according to standard works on this subject (cf. [Parker 1994]).

Suppose we have a potential scan consisting of a set of U_j values at a number of positions j above the substrate. These measurements are related to actual CPDs on the surface by the response function:

$$U_j = \sum_k h_{jk} \text{CPD}_k \quad (6.6)$$

Here h_{jk} is the probe response to a CPD at area k on the substrate, measured in point j above the substrate. It may seem obvious to determine the CPD_k values in areas right under the j positions. Then there is a simple one on one correspondence between the j and k areas and h_{jk} is a square matrix. Assuming that h_{jk} is somehow known, one can then in principle obtain the CPD_k values for the measured U_j by a simple matrix inversion. In practice however, this straightforward way of doing will not work for very profound reasons. In brief, these are:

- “Incompleteness” of the U_j data;
- CPD distributions on the surface that are “invisible” for the probe.

Incompleteness of the data means that the lateral steps of the scan are too large to describe the potential variations on the surface or that the measured U_j are also influenced by CPDs outside the scanned area. A third reason of incompleteness is noise. Due to incompleteness, the result of the inversion will contain incorrect contributions from CPD distributions that are not well “visible” for the probe. In the theory of inversion problems, such source distributions are generally referred to as “non-radiating sources”. Typically for scanning probes, non-radiating sources are CPD distributions with short-range lateral variations.

Several “smoothing techniques” are used to improve the solution of the inversion problem. In chapter 7, three such techniques are discussed and compared. This chapter handles a completely different problem, but the structure of the inversion problem is the same. For instance, one of these techniques could be implemented by determining the CPD_k values in fewer points than the number of U_j points. Then h_{jk} is no longer a square matrix and the inversion can be done by a least square fit. Additional a priori knowledge on the solution could then improve the solution, by which the inversion problem is turned into an optimization problem.

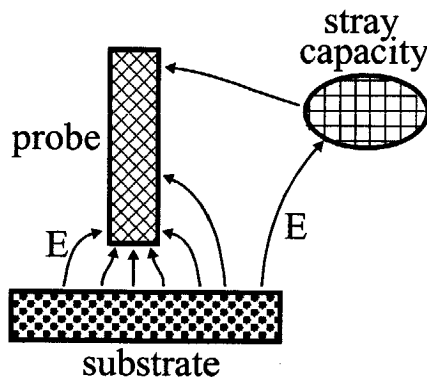


Figure 6.4. The SKP above a substrate with leakage of the electric field to a stray capacity.

6.2.2.2 Stray capacities

In the above model it was assumed that all electric field lines flow between the probe and the substrate. When this is not the case and a significant part of the electric field leaks out to objects in the environment, as illustrated by figure 6.4, the electric potentials of the objects in the environment also contribute to the measurement. Then, equation (6.4) for the charge on the probe can still be used:

$$Q_p = -\sum_k C_{pk} V_{pk}$$

But now the sum over k also contains contributions from the objects in the environment. The coupling between the probe and the objects in the environment are the mutual capacities and are called “stray capacities”.

One effect of stray capacities is that the measured CPD will depend on the distance between the probe and the substrate, even for a homogeneous substrate. The reason is that when the distance between the probe and the substrate is increased, the contribution of stray capacities to the measured CPD becomes more important.

Stray capacities can significantly bias the measured CPD and when one is interested in the actual value of the CPD, the setup requires special care to eliminate them. A solution is to use a guard electrode (cf. [De Boer 1973, Nabhan 1997]). In case one is only interested in relative differences of the CPD, it suffices to keep the contribution of stray capacities constant by maintaining a constant distance between the probe and the substrate during the measurement. This means that the SKP must be equipped with a distance control system.

6.2.3 Distance control

There are several reasons why the scanning Kelvin probe needs a proper distance control system:

- Preventing the probe from touching the substrate;
- Maximizing the probe current for more accurate potential measurement;
- Optimizing the lateral resolution;
- Keeping the influence of stray capacities constant.

Touching the surface should be prevented not only because it may damage the probe, but also because the probe may pick up material from the surface, which disturbs its reference potential. The other reasons for a distance control system are motivated by the above discussions. It follows that not only a constant, but also a close distance between the probe and the substrate should be maintained. In practice, one usually chooses the distance about equal to the dimensions of the probe tip, because closer distances no further improve the resolution and increase the risk of crashes. In the following subsections, a number of distance control techniques are discussed.

6.2.3.1 Distance control with the capacity between the probe and the substrate

Perhaps the most obvious way of distance control is to keep the capacity C_{ps} between the probe and the substrate constant during a scan. This quantity can be determined by an impedance measurement, for which a different frequency must be chosen than the frequency of the probe vibration, to prevent interference with the probe current. Nabhan for instance measured the capacity at 100kHz while the probe was vibrating at 20kHz [Nabhan 1997].

The capacity between the probe and the substrate is generally very small and not easy to measure. For instance, when the probe diameter and the scanning distance are in the range of 50 to 100 μ m, C_{ps} will be only 0.1 pF. In addition for accurate distance control, this capacity must be measured with a precision of about 1%. This already requires special equipment and for smaller distances and probe dimensions, C_{ps} becomes even smaller. In fact, when the geometry of the probe and the substrate is scaled by some factor K , the mutual capacity is scaled by the same factor. This follows from the fact that when for some geometry all distances and charges are multiplied by a factor K , the potentials on all objects remain the same. Hence when a probe with the same shape but 10 times smaller is used and the distance between the probe and the substrate is also 10 times smaller, C_{ps} will be 10 times smaller.

6.2.3.2 High U distance control

For the CPD measurement, the voltage U is only used as compensation voltage to null the probe current $I(t)$. However, U can also be used for an easily implemented distance control system. For the commercial SKP of UBM, this is done by keeping the amplitude of the probe current constant at a high U . This method is explained by equation (6.3) for the probe current, which reads in first order:

$$I(t) \cong V_{ps} \omega \delta z C_{ps}'(z_0) \cos(\omega t) \equiv \delta I \cos(\omega t)$$

The amplitude of $I(t)$ reads:

$$\delta I = V_{ps} \omega \delta z C_{ps}'(z_0) = (CPD - U) \omega \delta z C_{ps}'(z_0)$$

This is in fact the first harmonic of the probe current, which is measured by a LIA. Now when U is much higher than the CPD variations on the substrate, one has:

$$\delta I \cong U \omega \delta z C_{ps}'(z_0)$$

And since ω and δz are constant during the scan, it follows that δI at a high and constant U is proportional to $C_{ps}'(z_0)$. Provided that $C_{ps}'(z)$ is a monotonous function of z , this means that δI at large and constant U is a measure for the distance. For the commercial SKP of UBM for instance, $U = 8V$ is used. A drawback of this distance control system is that after all, it is somewhat sensitive to CPD variations on the surface. In addition, a high value of U may influence the electric charge distribution on the substrate.

6.2.3.3 Second harmonic distance control

An elegant method to overcome the dependence of the distance control system on lateral CPD variations is to use the signal distortion as a measure of the distance [Baumgärtner 1992, Mäckel 1993, Ließ 1997]. In section 6.2.1, it was shown that the probe current $I(t)$ is not a perfect sinusoidal wave, but contains a well-defined distortion, given by (6.3):

$$I(t) \cong V_{ps} \omega \delta z \left[C_{ps}'(z_0) \cos(\omega t) + \frac{\delta z}{2} C_{ps}''(z_0) \sin(2\omega t) \right]$$

Now when V_{ps} is not equal to zero (hence when U is not equal to the CPD), the ratio of the amplitudes of the first and second harmonics reads:

$$\frac{\delta I_2}{\delta I_1} = \frac{\delta z}{2} \times \frac{C_{ps}''(z_0)}{C_{ps}'(z_0)}$$

This ratio is dependent on the distance and independent of the CPD. The amplitudes δI_1 and δI_2 can often both be determined by a LIA. In principle, the value of U does not matter, but it is better not to take a too small U value, for accurate determination of δI_1 and δI_2 . (And with an accuracy independent of the lateral CPD variations.)

For using the ratio $\delta I_2/\delta I_1$ as a measure of the distance, it should be a monotonous function of the distance. This holds for the double plate geometry [Baumgärtner 1992], and for some other geometries for which an analytic expression for $C_{ps}(z)$ exists [Mäckel 1993]. In practice it is observed that for probes with a fine tip, $C_{ps}(z)$ is roughly proportional to z^{-1} around the operational distance, hence decreases more slowly with z than for the double plate geometry, for which $C_{ps}(z)$ is proportional to z^{-2} (cf. [Meijers 1996, Exalto 1998]). In general, when $C_{ps}(z)$ is proportional to $z^{-\alpha}$, $\delta I_2/\delta I_1$ is proportional to z^{-1} and can indeed be used for distance control.

6.2.3.4 Slope distance control

Another way of obtaining $C_{ps}'(z)$ is to measure the amplitude of the probe current at two different values of U :

$$\delta I(U_1) = (CPD - U_1) \omega \delta z C_{ps}'(z_0) \quad (6.7a)$$

$$\delta I(U_2) = (CPD - U_2) \omega \delta z C_{ps}'(z_0) \quad (6.7b)$$

The difference of the two reads:

$$\delta I(U_2) - \delta I(U_1) = (U_2 - U_1) \omega \delta z C_{ps}'(z_0)$$

This difference is also proportional to $C_{ps}'(z)$ but in contrast with δI , does not depend on the CPD. Hence when this quantity is used for distance control, one has no problems with lateral CPD variations on the substrate [Exalto 1998].

Also the CPD can be obtained from a measurement of δI at two different values of U . This is done by elimination of $C_{ps}'(z)$ from equations (6.7a) and (6.7b):

$$\text{CPD} = \frac{U_1 \delta I(U_2) - U_2 \delta I(U_1)}{\delta I(U_2) - \delta I(U_1)}$$

which provides another way of obtaining the CPD than by nulling the probe current. Instead of measuring δI for only two values of U , more accurate results are obtained when δI is measured for a larger number of U values and by fitting a straight line through the measurements, as shown in figure 6.5. The intersection with the $\delta I = 0$ axis then gives the CPD, while the slope of the line is proportional to $C_{ps}'(z)$ and can be used as a measure for the distance. Interestingly enough, determining U from a number of δI measurements for different U values is commonly known, but the idea of using the slope of the same graph as a measure of the distance seems to be new.

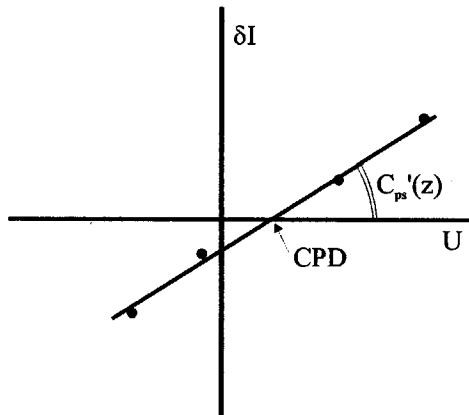


Figure 6.5. Simultaneous CPD and $C_{ps}'(z)$ measurement by fitting a straight line to a number of δI measurements for different values of U .

6.2.3.5 Other methods of distance control

The above distance control methods are all based on the capacity between the probe and the substrate. They are also all contactless methods of distance control, which has the advantage that the probe is not contaminated. In practice however, distance control is also done by touching the surface, which is for instance the case when the atomic force microscope AFM is used as an SKP (cf. [DeVecchio 1998, Kikukawa 1996, Schmutz 1998, Seminikhin 1997]). Other methods of distance control are possible by using a laser. This can be done in different ways, such as:

- Projecting a laser spot is on the substrate and observed it from an angle;
- Laser interferometry;
- Using an autofocus system, similar to CD-players.

An important difficulty with laser distance control methods is that it is not trivial how to bring the probe to exactly the area where the surface profile was measured. An option is to make very fine marks on the surface that can both be detected by the laser and the scanning probe. This will be possible but not very practical.

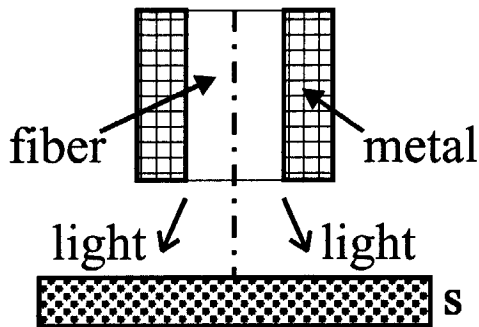


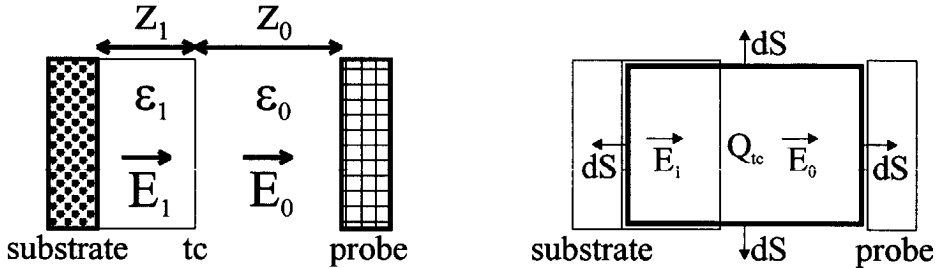
Figure 6.6. *Combined probe and waveguide. The core of the probe is an optical fiber through which a laser beam is guided.*

An idea to overcome these difficulties is to guide the laser beam through the core of the probe by using an optical fiber as the probe [Exalto 1998]. Since the fiber is not an electric conductor, a metal layer (or perhaps a conducting polymer) has to be applied on the fiber as shown in figure 6.6. With such a combined probe and waveguide, one can think of several ways to come to a laser distance control system:

- Measure the intensity of the light reflected on the surface and reentering the fiber. The amount of light returning in the fiber strongly depends on the height of the probe. The fiber then transports light in two directions and a beam splitter will be necessary to distinguish the reflected light. Since the probe vibrates, the reflected signal will contain a constant and a modulated component. By considering the ratio of these components, possible differences in the specific reflection on the surface can be eliminated;
- Etch a lens on the probe tip with focal length L and use an autofocus system to keep the probe at a constant height L , perhaps using the distance control system of CD-players;
- Use a double fiber as probe tip. One of the fibers can then be used to guide the laser beam to the surface, while the other is used to detect the reflected signal. In this way one does not need to worry about beam splitting techniques.

When the intensity of the reflected light is used, it will be better to use a multimode fiber, since in a monomode fiber, the reflected signal will not be propagated very well and may become too small to measure. Typically, the core of a multimode fiber has a diameter of

10 μm , which is then the minimum size of the probe tip. The most suited lasers for this job are probably GaAs-lasers. If desired, these can be easily modulated in contrast with HeNe-lasers.



Figures 6.7a and 6.7b. *The electric field between the probe and the substrate when an insulating coating is present and the surface S (denoted by a bold rectangle).*

6.2.4 The influence of insulating and poorly conducting coatings

In the previous sections, we only dealt with conducting substrates such as metals and electrolytes. In this section we will discuss the influence of coatings on Kelvin probe measurements. We start by considering the case of a perfectly insulating coating and later discuss the influence of a finite conductivity.

6.2.4.1 Perfectly insulating coatings

In figure 6.7a, the measurement of a substrate with an insulating coating is sketched. It will be assumed that the substrate, the coating and the probe are parallel. The coating has thickness z_1 and permittivity ϵ_1 , and "tc" indicates the top of the coating. The distance between the probe and the top of the coating is z_0 , and ϵ_0 is the permittivity of the air gap. The top of the coating may contain a homogeneously distributed charge Q_{tc} . This leads to a one-dimensional geometry, where only the z -component of the electric field is relevant.

Now in order to determine what the Kelvin probe measures, a relation between the charge Q_p on the probe and the potential difference V_{ps} between the probe and the substrate will be evaluated. For that purpose, we will apply Gauss's law:

$$\int_S \epsilon \vec{E} \cdot d\vec{S} = Q$$

over the closed surface S indicated by figure 6.7b. The surface encloses the charge Q_{tc} on top of the coating, hence the outcome of integral is:

$$-\epsilon_1 E_1 A + \epsilon_0 E_0 A = Q_{tc}$$

Here A is the area of the probe and the substrate. The contribution of E_1 is negative because E_1 is directed opposite to dS . For a really one-dimensional geometry, the area A and the charge Q_{tc} are infinite and therefore it is more natural to consider the charge density $\sigma_{tc} = Q_{tc}/A$ which is finite. With that, one can write:

$$-\epsilon_1 E_1 + \epsilon_0 E_0 = \sigma_{tc} \quad (6.8)$$

The charges on the substrate and the probe Q_s and Q_p are again proportional to the electric fields originating and disappearing in the substrate and the probe:

$$Q_s = \epsilon_1 E_1 A$$

$$Q_p = -\epsilon_0 E_0 A$$

In terms of surface charges, these expressions read:

$$\sigma_s = \epsilon_1 E_1 \quad (6.9a)$$

$$\sigma_p = -\epsilon_0 E_0 \quad (6.9b)$$

Substitution in (6.8) gives the following relation between all charge densities in the system:

$$-\sigma_s - \sigma_p = \sigma_{tc} \quad (6.10)$$

This implies that the sum of all charges in the system must be zero, consistent with the fact there is no leakage of electric field lines to the environment. Now the potential difference between the probe and the substrate is the sum of the potential differences over the coating and the air gap:

$$V_{ps} = z_1 E_1 + z_0 E_0$$

Substitution of (6.9a) and (6.9b) for the electric fields gives:

$$V_{ps} = \frac{z_1}{\epsilon_1} \sigma_s - \frac{z_0}{\epsilon_0} \sigma_p$$

Finally σ_s can be eliminated by equation (6.10), which leads to:

$$V_{ps} = -\frac{z_1}{\epsilon_1} \sigma_{tc} - \left(\frac{z_0}{\epsilon_0} + \frac{z_1}{\epsilon_1} \right) \sigma_p$$

This is the relation we were looking for, a relation between the charge on the probe and the voltage between the probe and the substrate. From this expression, it can be evaluated what the Kelvin probe measures. The charge on the probe reads:

$$\sigma_p = -\frac{V_{ps} + z_1 \sigma_{tc} / \epsilon_1}{z_0 / \epsilon_0 + z_1 / \epsilon_1} \quad (6.11)$$

When the probe is vibrated, z_0 is modulated and the alternating current in the circuit is measured. Now for the potential measurement, U is adjusted such that the amplitude of the probe current is zero, hence such that the variation of σ_p is zero. From equation (6.11) it follows that σ_p is always zero, independent of z_0 , when:

$$V_{ps} + \frac{\epsilon_1 \sigma_{tc}}{z_1} = 0$$

which corresponds to the following compensation voltage:

$$U = \text{CPD} + \frac{\epsilon_1 \sigma_{tc}}{z_1} \quad (6.12)$$

This expression shows how much the measured potential is influenced by a surface charge σ_{tc} on top of an insulating coating with thickness z_1 and permittivity ϵ_1 . It follows that an uncharged, insulating coating does not contribute to the CPD, but this analysis does not include the surface dipole of the polymer and the chemical adhesion of the coating to the metal, which also may influence the measured CPD.

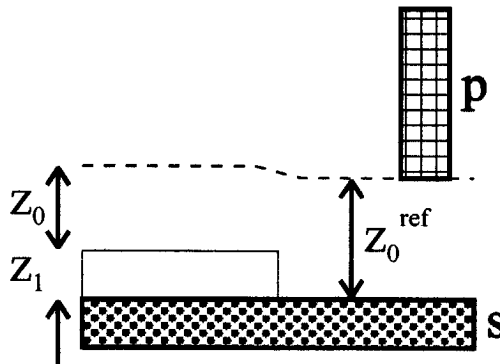


Figure 6.8. The influence of the thickness of the coating on distance control. Shown is a scan over a partly coated substrate.

6.2.4.2 The influence of an insulating coating on distance control

We will now also evaluate the effect of an insulating coating on the capacitive distance control systems of section 6.2.3. For the method of 6.2.3.1, the capacity between the probe

and the substrate is measured. This total capacity is equal to the capacity of the coating C_1 and the air gap C_0 in series:

$$\frac{1}{C_{ps}} = \frac{1}{C_1} + \frac{1}{C_0} = \frac{z_1}{\epsilon_1 A} + \frac{z_0}{\epsilon_0 A}$$

It follows that when C_{ps} is kept constant during distance control and when the coating thickness z_1 is constant over whole surface, the distance z_0 between the probe and the top of the coating is kept constant, which is all right.

It also follows that variations of the coating thickness will disturb the distance control system. Suppose for instance that on part of the surface a coating has thickness z_1 while the coating is absent on another part, as shown in figure 6.8. Now the distance control system will keep C_{ps} constant, which means that:

$$\frac{z_0}{\epsilon_0} + \frac{z_1}{\epsilon_1} = \frac{z_0^{ref}}{\epsilon_0}$$

where z_0^{ref} is the distance between the probe and the substrate at the area where the coating is absent. From the latter equation, the distance z_0 between the probe and the top of the coating follows:

$$z_0 = z_0^{ref} - \frac{\epsilon_0}{\epsilon_1} z_1$$

Hence the distance between the probe and the top of the substrate is smaller when the coating is thicker and above a certain value of z_1 , the probe will even touch the surface, with all possible consequences.

For the distance control system of 6.2.3.2 and 6.2.3.4, an expression for the probe current for arbitrary U is needed, which is obtained by:

$$I = -\frac{dQ_p}{dt} = -A \frac{d\sigma_p}{dt} = -A \frac{d\sigma_p}{dz_0} \frac{dz_0}{dt}$$

With that and with the equation (6.11) for σ_p , one can derive the following for the amplitude of the probe current:

$$\delta I \propto \frac{d\sigma_p}{dz_0} = \frac{[V_{ps} + z_1 \sigma_{tc} / \epsilon_1]}{[z_0 / \epsilon_0 + z_1 / \epsilon_1]^2} \times \frac{1}{\epsilon_0}$$

This is the quantity kept constant by the distance control systems of 6.2.3.2 and 6.2.3.4. We see that again $z_0/\epsilon_0 + z_1/\epsilon_1$ is kept constant, similar to the distance control system of 6.2.3.1, hence the same conclusions hold. (Although for the system of 6.2.3.2 it is also required that the lateral variations of the CPD and the influence of static charges can be discarded with respect to U .)

Finally for the distance control system of 6.2.3.3, the ratio $C_{ps}'(z_0)/C_{ps}''(z_0)$ is kept constant, which in the end also comes down to keeping $z_0/\epsilon_0 + z_1/\epsilon_1$ constant. Without considering static charges, J. Ren studied the distance control system of 6.2.3.2 and came to the same conclusions [Ren 1995].

6.2.4.3 The influence of the coating conductivity

When the coating has a finite conductivity, any voltage difference in the coating is unstable and will be leveled out in thermodynamic equilibrium. However for many coatings, it can take hours or days before the thermodynamic equilibrium is actually reached. When this is the case, the coating can for our purposes be regarded as a perfect insulator and the analysis of the previous subsection is sufficient.

The speed of charge transport in a material is characterized by the time constant τ , which is the ratio of the permittivity ϵ and the conductivity γ . The time constant appears in all kinds of charge transport problems in materials, for instance for the case where a charge $Q(0)$ on top of a coating flows away to the metal substrate underneath. This case is equivalent to the problem of a capacitor C_1 that is discharged through a resistor R_1 . In that case, the charge decays exponentially by:

$$Q(t) = Q(0) \exp[-t/\tau_1]$$

where $\tau_1 = R_1 C_1$ is the time constant. Now by substitution of $C_1 = \epsilon_1 A / z_1$ and $R_1 = z_1 / \gamma_1 A$, where A is the area of the capacitor plates, we get $\tau_1 = R_1 C_1 = \epsilon_1 / \gamma_1$, which is indeed the earlier mentioned ratio of the permittivity and the conductivity.

Now the time constant τ_1 must be compared to the duration of the measurements T_{meas} and the period of the probe vibration $T_{\text{vib}} = 2\pi/\omega$. This leads to the definition of the following three regions for τ_1 :

1. $\tau_1 < T_{\text{vib}}$. The coating behaves as a perfect conductor;
2. $T_{\text{vib}} < \tau_1 < T_{\text{meas}}$. The coating will reach a thermodynamic equilibrium in which all potential differences are leveled out and no static charges remain. However the charge transport in the coating is too slow to follow the probe vibrations;
3. $\tau_1 > T_{\text{meas}}$. In this case, the coating behaves as a perfect insulator. Any present charges will be static.

Apart from the earlier mentioned surface dipole effects and the chemical adhesion of the coating, the coating will have no effect on the measured CPD, except when τ_1 is in region 3 in which case static charges can be present. The effect on the capacitive distance control system follows from the consideration that in region 1, the coating is able to follow the probe vibrations and the distance control system will keep the distance between the probe and the top of the coating constant. In region 2, the coating does not follow the probe vibrations and the distance control system works similar as in region 3, although static charges will not appear when τ_1 is in region 2.

It must be recalled that our whole analysis of coated substrates is carried out in one dimension and the considerations on the capacitive distance control system may very well be different when the actual three-dimensional geometry is considered, such as by a finite element analysis. Another question that is also not answered by a one-dimensional analysis, is whether lateral CPD differences underneath coatings can be detected. Again, this should actually be studied by finite element analysis, but one may in this case expect that when τ_1 is in region 1 or 2, the potential differences in the coating will be leveled out rapidly enough such that the CPD differences underneath the coating will not be detected.

6.3 Surface potential measurements

A number of measurements were done with a homemade SKP described elsewhere [Meijers 1996, Exalto 1998]. The probe was a stainless steel needle, which vibrated with a frequency of 1000 Hz using a loudspeaker. A number of standard tests are presented, followed by a number of measurements on characteristic systems such as metals with an electrolyte layers and coated substrates.

6.3.1 Standard tests on metal surfaces

Here a number of standard tests are discussed such as the stability of the measured potential and scans of surfaces with well-defined lateral potential distributions. Also several distance control systems will be compared.

6.3.1.1 Stability of the measured potential

Non-reproducible drifting of the measured CPDs was regularly observed, sometimes within hours. On one and the same substrate, the measured potential then varied by up to 200mV. Drifting of the measured CPD is a well-known phenomenon for the Kelvin probe and in fact for the commercial SKP of UBM, the user is supposed to regularly gauge the probe by a measurement on a reference sample.

Drifting of the measured CPD may be due to variation of the temperature in the electric circuit, as already mentioned in section 6.2.1. This explanation is supported by the fact that drifting was often observed after the setup was switched on, hence when the electronics was heating up. Another reason of measuring a different CPD can be that the metal composition of the probe tip has changed, which may be due to an accidental contact with a substrate.

6.3.1.2 Measurement of a well-defined voltage

To verify if the apparatus measures potentials on surfaces, a test sample was made from the carbon film of a potentiometer. The area of the carbon film was 1mm wide and 9mm long and its total resistance was 10k Ω . Now by using a 1.5V battery, a stable and linearly

increasing voltage was put along the carbon film. In figure 6.9, a potential scan along 6mm of the carbon film is shown. We see that over this distance, the potential increases by about 1V, in agreement with 1.5V over 9mm. Note that in the middle of the carbon film, an anomaly appears, which was also visible when the battery was disconnected. This suggests an inhomogeneity of the carbon film.

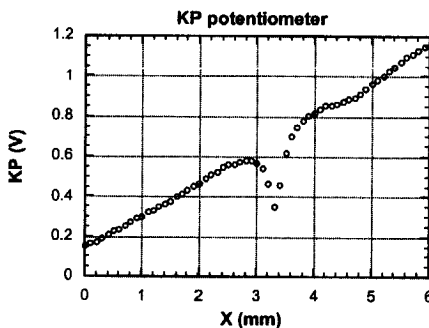


Figure 6.9. Potential scan over the carbon film of a potentiometer over which a direct voltage is put.

6.3.1.3 Measurement of the resolution

For the SKP, two kinds of resolution are important, the potential resolution and the lateral resolution. These resolutions are not independent since the lateral resolution is improved by a finer probe tip, which at the same time leads to a smaller probe current and hence, to a lower potential resolution.

For measuring the lateral resolution, a sample with a sharp lateral variation of the CPD is needed. For that purpose, a copper line was sputtered through a mask on an aluminum sample. The width of the copper line was 1mm and its thickness less than $1\mu\text{m}$. When going from aluminum to copper, the electric potential makes a jump equal to the CPD between aluminum and copper and the sharpness of the jump is a measure for the lateral resolution.

In figure 6.10, potential scans over the copper-aluminum sample are shown, recorded at four different scanning distances, which were 10, 20, 100 and $200\mu\text{m}$. The diameter of the probe tip was about $100\mu\text{m}$. We see that below $20\mu\text{m}$, the resolution is basically determined by the size of the probe tip. The potential resolution is given by the statistical errors in the potential scans and about 5mV.

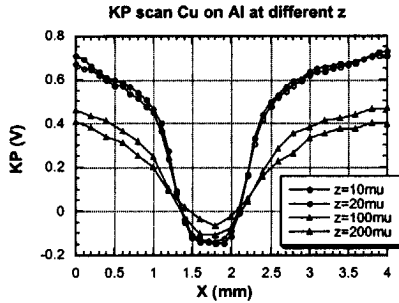


Figure 6.10. Potential scans over a copper line on aluminum at different heights, showing the influence of the scanning distance on the resolution.

6.3.1.4 The influence of stray capacities

In subsection 6.2.2.2, it was discussed that the influence of stray capacities shows up as a dependence of the measured CPD on the distance z between the probe and the substrate. In figure 6.11, the effect is shown by a CPD measurement on a homogeneous platinum surface as a function of z . It follows that for our setup, the measured CPD varies by about 5mV per 10µm. Note that variations of the measured potential as a function of height can also be due to accidental inhomogeneities on the surface. This can be verified by checking whether the z -dependence of the CPD is the same at different positions on the surface.

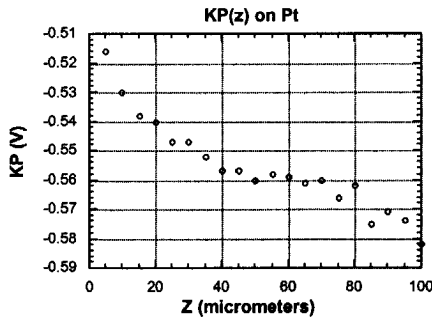


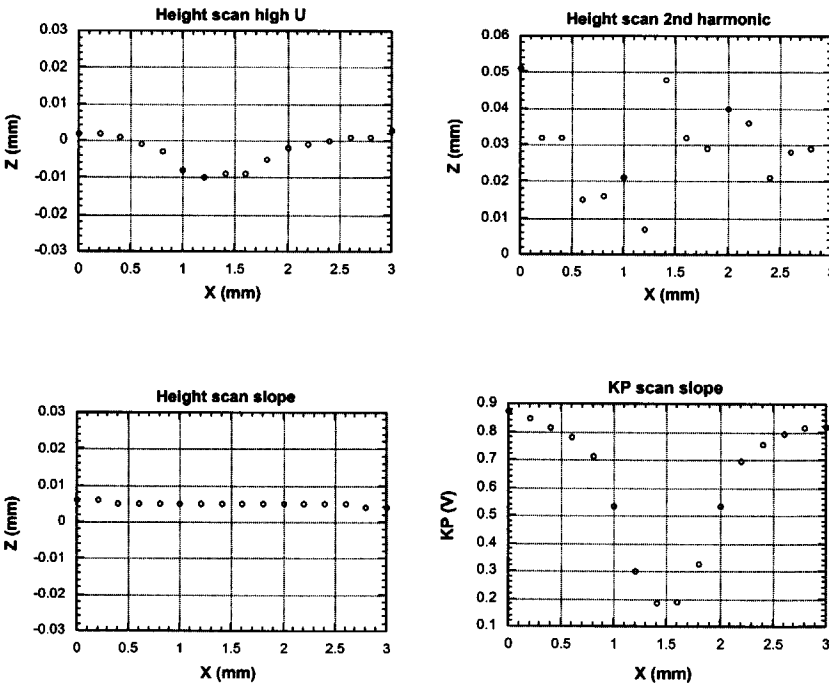
Figure 6.11. Influence of the distance on the measured potential on a homogeneous platinum sample, showing the influence of stray capacities.

6.3.1.5 Distance control

Three methods for capacitive distance control were implemented and their performances were compared:

- High U distance control;
- Second harmonic distance control;
- Slope distance control

These are the distance control systems of sections 6.2.3.2, 6.2.3.3 and 6.2.3.4 respectively. The distance control systems were compared by potential scans on the aluminum sample with a copper line, earlier mentioned in section 6.3.1.3. The probe was again a needle with a tip diameter of $100\mu\text{m}$ and the scanning distance was $50\mu\text{m}$.

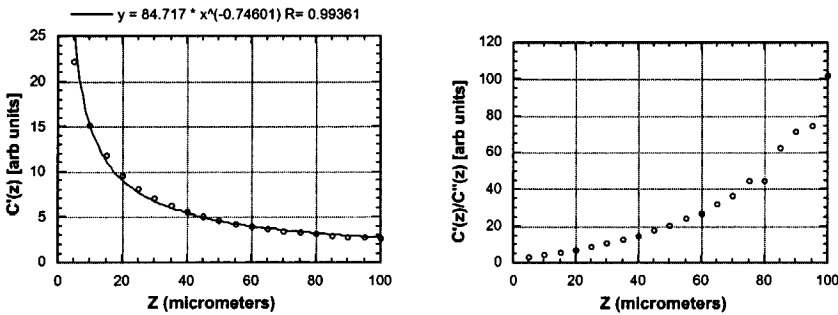


Figures 6.12a, 6.12b, 6.12c and 6.12d. Height scans with three different distance control systems over the copper-aluminum sample. The potential scan is given in the last figure. The scanning height was $50\mu\text{m}$.

The height scans for the three distance control systems are presented in figures 6.12a, 6.12b and 6.12c. The corresponding CPD scan is given in figure 6.12d. For high U distance control, $U = 8\text{V}$ was used, as for the commercial apparatus of UBM. For second harmonic distance control, $U = 2\text{V}$ was used and for slope distance control, the sequence $U = -2\text{V}, -1\text{V}, 1\text{V}, 2\text{V}$ was used. In these figures, only the differences in the z-values are relevant.

It follows that for high U distance control, the CPD variations on the surface cause a systematic error of about $10\mu\text{m}$. This is explained by the discussion in section 6.2.3.2. The second harmonic distance control system does not give a systematic error, but the statistic errors are large. The slope distance control system gives the best results, the copper line does not cause a systematic error and the statistical errors do not exceed the minimum step size of the motion controller, which is $1\mu\text{m}$.

The high U and slope distance control systems are based on $C'(z)$, while the second harmonic distance control system is based on the ratio $C'(z)/C''(z)$. These quantities were measured for the probe tip with a diameter of $100\mu\text{m}$ and are plotted in figures 6.13a and 6.13b. In good approximation, $C'(z)$ is proportional to $z^{-0.75}$. In section 6.2.3.3 it was mentioned that $C'(z)/C''(z)$ is proportional to z when $C'(z)$ is a power of z and figure 6.13b shows that for z below $60\mu\text{m}$ this is indeed roughly the case.



Figures 6.13a and 6.13b. The z -derivative of the mutual capacity between the probe and the substrate and the ratio of the first and the second derivative. A power is fitted through the measurement of the z -derivative.

A good heuristic relation for $C'(z)$ or $C'(z)/C''(z)$ is useful for adjusting the height during distance control. For all probe tips used it was found that $C'(z)$ was rather well approximated by a negative power of z . This power was always roughly around -1 . We have no explanation for this behavior and analytical expressions for the capacity between the probe and the substrate do not exist for most probe geometries.

For the double plate geometry, $C'(z)$ is proportional to z^{-2} which disagrees with the observed z^{-1} behavior. Another geometry for which an analytic expression for the capacity exists is a spherical probe above a plate (cf. [Mende 1974, Ederle 1990, Mäckel 1992]). The capacity of this geometry has the form $C(z) = r \times f(z/r)$, where r is the radius of the sphere and $f(z/r)$ is a function of the ratio z/r . This follows from the fact the mutual capacities of a system scale with a factor K when the whole geometry is scaled by a factor K , as already mentioned in section 6.2.3.1. By consequence, dC/dz is a function of the dimensionless ratio z/r . (This also holds for other geometries such as a cone above a surface or a cone above a sphere and so on.)

In figure 6.14, dC/dz of a spherical probe above a plate is plotted as a function of z/r . It follows that for $z \gg r$, dC/dz is proportional to z^{-2} , while for $z \ll r/10$, dC/dz does not change much with z . Around the operational distance however, where z is somewhat smaller than r , dC/dz is roughly proportional to z^{-1} , which gives some support for the observation that around the operational distance, $C'(z)$ is inversely proportional to z .

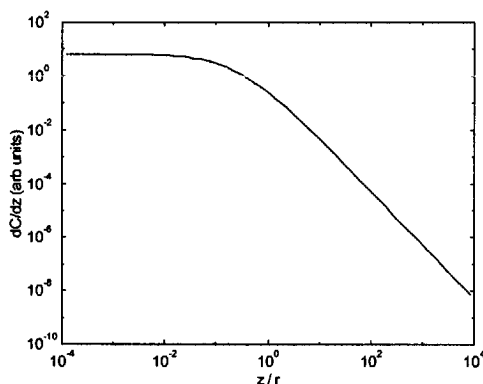


Fig. 6.14. Derivative of the capacity between a sphere and a plate as a function of the ratio of the mutual distance z and the radius of the sphere r .

6.3.2 Metal-electrolyte systems

The influence of thin electrolyte layers on the measured potential can be easily demonstrated by blowing humid air over a clean metal surface during a series of KP measurements. For instance, a potential drop of 20mV was observed when this was done with polished aluminum.

For better-defined measurements on metal-electrolyte systems, thicker metal plates were used in which a hole with a diameter of 1 cm and depth of 1mm was milled. The hole was filled by electrolyte, as shown in figure 6.15 and the potentials V_m and V_e above the metal and the electrolyte were measured with respect to the probe. As discussed in chapter 5, the potential difference $V_m - V_e$ is equal to the corrosion or mixed voltage plus a constant offset value due to the different surface dipoles of the metal and the electrolyte. Lateral resolution was not important for these measurements and we used a larger probe with a tip diameter of about 1mm.

Two typical and relatively simple systems were chosen, which are CuSO_4 in water on a copper surface and H_2SO_4 in water on an iron surface. In the first system, a thermodynamic equilibrium is expected, while the second system will corrode.

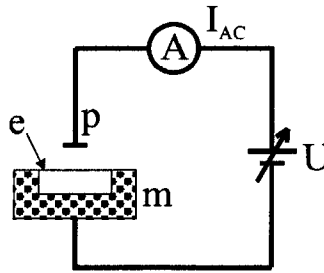


Figure 6.15. Measurements on metal-electrolyte systems. A hole is milled in the metal surface (denoted by *m*) and filled with electrolyte (denoted by *e*).

6.3.2.1 Copper in a CuSO_4 solution in water

For the commercial apparatus of UBM, the electric potential above CuSO_4 in water on copper is used as a reference to correct for possible drifting of the probe potential. Assuming that the system is in thermodynamic equilibrium, the voltage over the copper-electrolyte interface is given by the Nernst equation (see chapter 5):

$$V_m - V_e = (V_m - V_e)^{\text{ref}} + \frac{k_B T}{2|q_e|} \ln \left[\frac{n_{\text{Cu}^{2+}}}{n_{\text{Cu}^{2+}}^{\text{ref}}} \right]$$

It follows that at room temperature, the voltage increases by 30mV when the Cu^{2+} concentration increases by a factor 10, hence when this system is used as a reference, one should always use the same CuSO_4 concentration. The influence of the Cu^{2+} concentration on $V_m - V_e$ was verified by working with a saturated CuSO_4 solution that was diluted 10, 100 and 1000 times. These solutions will be referred to D1, D2 and D3 (adopting standard homeopathic notations). We did not use the saturated CuSO_4 solution because of the rapid precipitation of CuSO_4 , which interrupts the electric contact between the copper and the electrolyte.

The results of these measurements were not very promising. For the D1, D2 and D3 solutions, the potential was between 160 to 190mV higher above the electrolyte than above the copper, corresponding to $V_m - V_e = -160$ to -190mV . Hence, the expected 30mV per factor 10 of dilution was not observed. The measurements indicate that $V_m - V_e$ is not only determined by the equilibrium of Cu^{2+} ions but also other reactions, such as for instance the reaction between oxygen and water.

6.3.2.2 Iron in diluted sulfuric acid

In this case, two different basins were filled with sulfuric acid of pH = 1 and pH = 3 and the potentials above the iron and above the two basins were measured. The resulting corrosion voltages were $V_m - V_e = -580\text{mV}$ for the pH = 1 site and $V_m - V_e = -710\text{mV}$ for the pH = 3

site. These values were measured about 5 minutes after immersion and did not change significantly for at least 30 minutes.

Following Stratmann, the measurements were verified by comparison with similar measurements in an electrochemical cell [Stratmann 1990]. For these measurements, we used a saturated AgSO_4 electrode. The results were -940mV for iron in sulfuric acid of $\text{pH} = 1$ and -1095mV for iron in sulfuric acid of $\text{pH} = 3$. The difference between these two voltages roughly agrees with the Kelvin probe measurements.

In addition the potential above diluted sulfuric acid of $\text{pH} = 4$ on iron was followed during one hour. From measurements in an electrochemical cell we knew that in this time interval, the potential of iron with respect to the AgSO_4 electrode decreases from about -950mV to -1090mV . With the Kelvin probe, a gradual decrease of $V_m - V_e$ from -520mV to -700mV was measured in this time interval, which roughly agrees with the measurement of the electrochemical cell.

Also note that the corrosion potential of iron in diluted sulfuric acid increases from $\text{pH} = 1$ to $\text{pH} = 3$, but is initially again lower for $\text{pH} = 4$. The corrosion rate however rapidly increases when the pH is lowered, which supports the conclusion of chapter 5 that the corrosion potential is no good measure for the corrosion rate.

6.3.3 Coated substrates

In chapter 5, it was argued that an insulating coating could influence a CPD scan in several ways:

- By the surface dipole of the coating;
- By the adhesion of the coating to the substrate underneath;
- By static charges.

Especially the second effect is interesting for us, since it opens perspectives for measuring the local adhesion of polymer coatings on metals in general and of PET coated ECCS after a DWI process in particular. Apart from PET coated ECCS, we also studied a number of other cases of insulating layers and coatings on metals. In fact we observed that about any material such as grease, ink or salt products affects the CPD on a substrate, but here we will limit ourselves to the more reproducible examples where a well-defined layer of a well-defined thickness could be applied.

6.3.3.1 PET coated ECCS

A major problem for measuring on PET coated ECCS are static charges. At room temperature, the electric conductivity of PET is so extremely low (see also chapter 2) that static charges do not flow away automatically, even in days. On the other hand, static charges are applied easily by friction such as rubbing by a tissue or even by an air stream. Hence in order to measure local adhesion differences of PET coated ECCS, the samples first need a pretreatment by which the static charges are removed. This can be done in the following ways:

- Heating the samples above the glass temperature of PET;
- A treatment by an anti-static liquid.

Of course, one wishes not to modify the PET when removing the static charges and therefore we choose for the second option with ethanol as anti-static liquid. Ethanol is well suited, since it wets PET very well and does not react with it.

For the anti-static treatment of PET with ethanol, it was found very important that the ethanol evaporated uniformly. This was achieved by holding the sample under an angle of 45° while ethanol was sprayed over it and while the sample was drying. The effect of inhomogeneous evaporation of ethanol is shown in figure 6.16. This figure shows a surface CPD scan of a sample on which an ethanol drop of a few mm evaporated. The figure shows that as the ethanol drop evaporating and becomes smaller, it concentrates charges on top of the PET coating, which is visible as a significant contribution to the local CPD of about 20V.

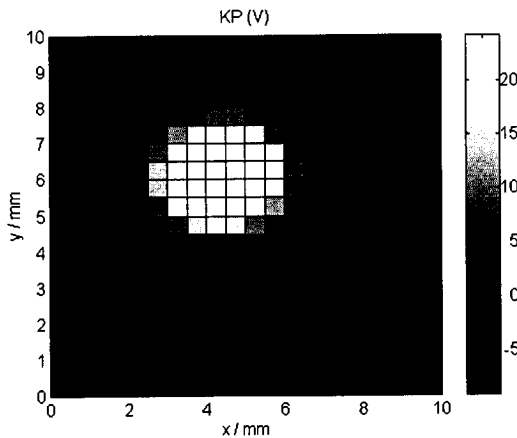
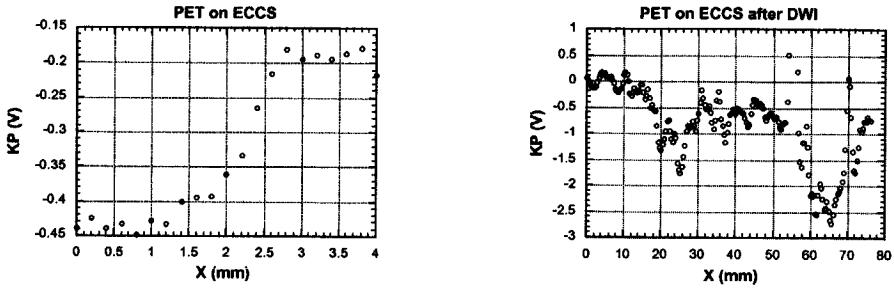


Figure 6.16. CPD scan over PET coated ECCS after an ethanol treatment. At the white spot, a large drop of ethanol had evaporated.

Since the undeformed PET coated ECCS samples had a somewhat rough top layer, it was difficult to remove the static charges by ethanol. Nevertheless, some results were obtained. Figure 6.17a shows a CPD scan over an ECCS sample that was partly coated with PET. Clearly, the CPD is influenced by the PET and decreased by about 0.25V. This difference can either be due to the surface dipole of PET or to the adhesion of PET on ECCS, hence it is not completely sure that this figure shows the adhesion voltage between PET and ECCS.



Figures 6.17a and 6.17b. CPD scans of a plane ECCS sample, partly coated with PET (the left part is coated) and along the height of a can fragment after a DWI process.

Figure 6.17b shows a CPD scan of a PET coated ECCS sample after deformation by a DWI process. The figure shows a scan over the height of a can wall and therefore the degree of deformation varies along the X-axis, as argued in chapter 4. After DWI, the PET coating looked considerably smoother and indeed the ethanol evaporated more homogeneously. The measurements reproduced rather well, including the dip of the CPD between 20 and 26mm and the two sharper dips at 20 and 26mm. These positions correspond to the larger radius of the can after the first deep drawing step process, which also left a visible mark. Above 30mm, the reliability of the CPD scan becomes worse with non-reproducible peaks that are probably due to remaining static charges. Nevertheless the global shape of the CPD scan with the reduction of about 2V was always observed.

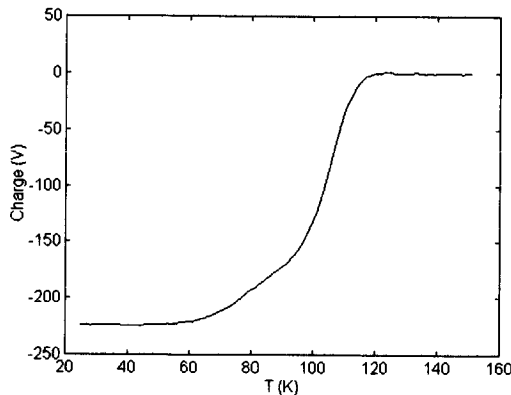
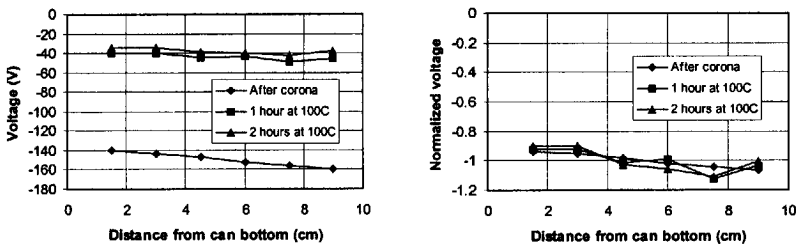


Figure 6.18. Discharge curve of PET on ECCS as a function of temperature. The temperature increased by 1°C per minute.

At room temperature, charges on PET are static as mentioned above and therefore, PET must be heated in order to obtain significant charge decay. In figure 6.18, the discharge curve of a PET coating of $25\mu\text{m}$ is shown as a function of the temperature, which increased by 1°C per minute. It follows that around 100°C , the charge flows away at a significant rate. This temperature is above the glass temperature of about 75°C , but below the crystallization range between 110 and 130°C , hence the PET coating can probably be heated up to 100°C without problems. We now did the following experiment for determining local adhesion differences:

- Homogeneously charge the coating and make a potential scan;
- Heat the coating up to 100°C for 1 hour and make a following scan;
- Heat the coating for another hour at 100°C and make a final scan.

In this way, potential scans were made of the same sample as used for figure 6.17b. The Kelvin probe of the corona apparatus was not equipped with a scanning facility and therefore the substrate was moved by hand, leading to a low lateral resolution. The results are shown in figure 6.19a and 6.19b. The potential scans show a more or less constant slope, which is already visible for the first potential scan that was made rapidly after the corona treatment. We see that the sample was not homogeneously charged, which is probably caused by an inclination between the (curved) sample and the grid electrode. In figure 6.19b, the measurements of figure 6.19a are normalized to the average potential values, which more clearly shows the local potential differences.

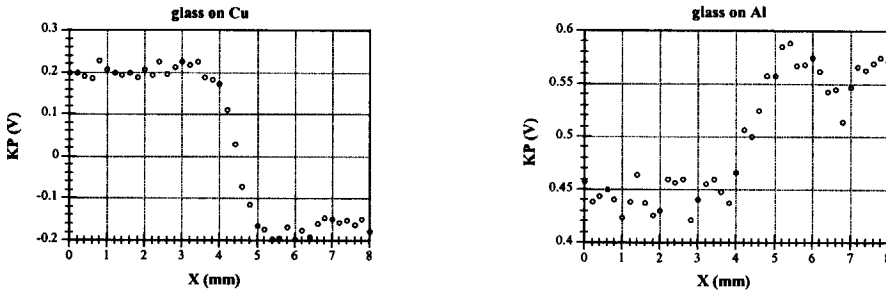


Figures 6.19a and 6.19b. *Electric potential scans along the height of a can fragment of PET coated ECCS after a DWI process. The sample was initially charged by the corona method. In the right figure, the potentials are normalized to the averages.*

The above results do not convincingly prove that the SKP can be used for measuring local adhesion differences. The CPD differences shown by the figures 6.17a-b and 6.19a-b could also be due to insufficient removal of static charges by ethanol in the first case, and by inhomogeneous initial charging by the corona method in the second case. It is clear that additional research with improved equipment is necessary for more decisive conclusions.

6.3.3.2 Glass on metals

Also some measurements were done with glass slides used for microscopes. Slides with a thickness of $150\mu\text{m}$ were used. The glass slides are plane and rigid and can therefore be used to cover a metal by an insulator with a well-defined thickness. First a glass plate was put on the aluminum sample with a copper line and a potential scan was made. To our surprise, no lateral potential differences were measured. This is not explained by the large scanning distance, because the distance between the probe and the underlying metal was only $200\mu\text{m}$. This suggests that for glass, the conductivity and the amount of mobile charge carriers are sufficient for leveling the potential differences.



Figures 6.20a and 6.20b. *Potential scans of glass on copper and aluminum. The scans start above the glass and end on the metal.*

In addition, separate potential scans of glass on copper and glass on aluminum were made and the results are shown in figures 6.20a and 6.20b. The scans start above the glass and end above the metal. The results are remarkable and not so easily explained. When considered as an insulator, glass could influence the potential scan by:

- The surface dipole;
- Adhesion to the metal;
- Static charges.

However since the potential scan above glass is rather smooth and well reproducible, static charges can be excluded. And since the glass plate was only placed on the metal, an adhesion effect is also not likely. Finally the surface dipole may contribute to the measured potential, but when this is the only contribution, one would expect the potential to differ from the metal by a fixed amount. That is to say, one would for instance expect the potential to be 0.2 V higher for both glass on copper and for glass on aluminum. However figures 6.20a and 6.20b show that in case of copper, the potential is increased by the glass while in case of aluminum, the potential is decreased by the glass.

A possible explanation is that the electronic conductivity of the glass slides is so high that a CPD between the glass and the metal is build up. Of course, the DC conductivity of glass is not really very high. The resistance of the used glass slides was more than $10\text{M}\Omega$

(along the thickness of the glass slide), corresponding to an electronic conductivity lower than $10^{-10}\Omega^{-1}\text{cm}^{-1}$. However, the associated time constant ϵ/γ can then still be in the order of milliseconds, which is low enough to equalize potential differences more rapidly than the measurement time mentioned in section 6.2.4.3. Note that the potential of glass on copper and aluminum still differs significantly by 0.25V, hence glass does not completely behave as a conductor in the sense that it has its own CPD.

In figure 6.21 a height scan is shown, where the probe moved from the glass slide on copper to the bare copper. We see that a step of about $150\mu\text{m}$ is taken when leaving the glass. (The Z value increases, because the motion controller was mounted upside down.) Hence the top of the coating is followed, which also indicates that the glass behaves as a conductor.

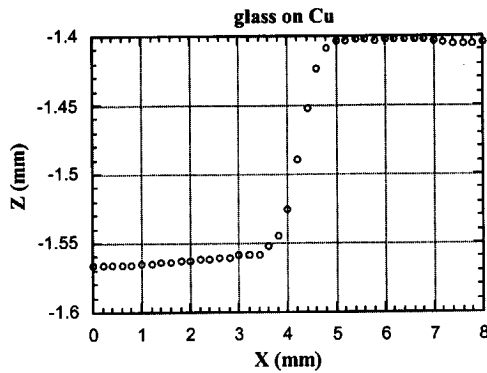


Figure 6.21. Height scan on a copper sample with a microscope slide, corresponding to the measurement of figure 6.20a.

6.3.3.3 Scotch tape[®] on metals

Scotch tape[®] has a well-defined thickness of $60\mu\text{m}$ and is easily applied on materials. The same experiments were done as with the glass slides, but led to different results. First when the tape was applied on polished copper or aluminum, the potential was 150mV higher above the tape in both cases. This implies that the contribution of the tape to the measured potential is basically due to the adhesion between the tape and the metal together with the surface dipole effect. Hence the Scotch tape[®] seems to behave as an insulator. Static charges were not so much a problem as with PET. Rubbing with a tissue for instance did induce charges on top of the tape but, as verified by the SKP, these disappeared within a minute.

We also applied Scotch tape[®] on the earlier mentioned aluminum sample with a copper line and at first sight, the result was perfectly in line with the idea that the tape behaves as an insulator, since a potential scan clearly showed the copper line. Hence potential differences were visible through the tape. We also tried to measure the influence of adhesion of the tape on the measured potential by not pressing the tape on the metal on part of the sample. This led to an unexpected result, which is shown in figure 6.27. Instead of an influence due to an adhesion voltage, we see that the copper line is no longer visible when the tape is not properly fixed. Now why would this be the case? A possible explanation is that

only the glue layer of the tape has some conductivity and that the charge carriers in the glue are neutralized when the tape is firmly attached to the metal.

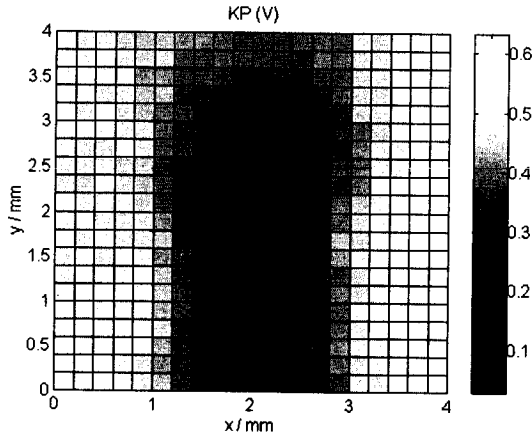


Figure 6.22. Aluminum with a copper line of 1 mm thickness, covered by Scotch tape[®]. Only at the lower part of the figure, the tape is properly fixed.

In fact it is interesting that potential differences are visible at all underneath Scotch tape[®], since the tape clearly shows some electric conductivity. (As mentioned above, the time constant of induced charges is about 1 minute). The reason is probably that the density of charge carriers in the tape is too low to produce the charge distribution required for leveling the potential differences.

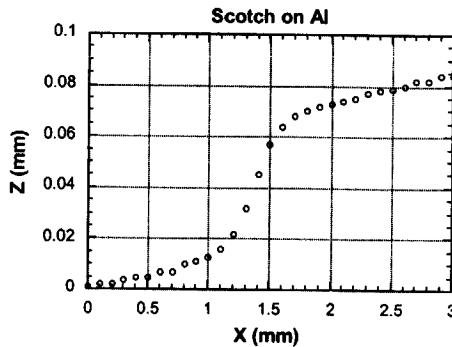


Figure 6.23. Distance control on an aluminum sample with Scotch tape[®]. The scan starts above the tape and ends on the metal.

On the other hand, the time constant of Scotch tape[®] is so low that the charge carriers in the tape, if any, are unable to follow the probe vibrations. This is illustrated by figure 6.23, where a height scan is shown of an aluminum sample that was partly covered by Scotch tape[®]. By elimination of the surface inclination, it follows that the probe is lifted by about 45 μm above the tape, which is 15 μm less than the thickness of the tape. Following section 6.2.4, this means that the tape qualifies for region 2.

6.3.3.4 Underpaint corrosion

One of the main motivations for the introduction of the SKP in corrosion research is to study underpaint corrosion. A particularly interesting case is filiform corrosion of painted aluminum (cf. [Lenderink 1995, Spoelstra 1999]). For this particular type of corrosion, delamination of the coating takes place in the form of a one-dimensional filament, while the coating remains intact. Now since the head of an active filiform is anodic while the tail is cathodic, filiform corrosion provides an almost ideal test case for an SKP scan (cf. [Lauterbach 1999]).

From the considerations of chapter 5, it follows that the result of a potential scan over an active filiform strongly depends on the conductivity of the electrolyte between the anodic head and the cathodic tail. For a high conductivity, and hence a high corrosion rate and a very active filiform, the potential over the whole filiform will be more or less constant and equal to the corrosion potential. The potential above the filiform then differs from the potential around the filiform by the corrosion voltage and also by a possible adhesion voltage between the coating and the aluminum. (The latter contribution is absent at the filiform where the coating has delaminated.) For a very low conductivity between the head and the tail, on the other hand, hence for a very slowly growing filiform, the potential differences between the head and the tail will be more clearly visible. One could therefore conclude that a flatter potential scan denotes rapid filiform corrosion. (However, one must be careful since a flat potential scan can of course also denote uniform corrosion or no corrosion at all).

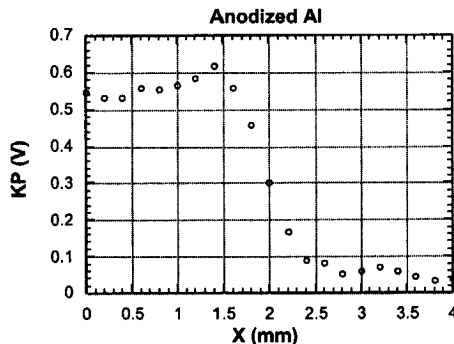


Figure 6.24. Potential scan over a partly anodized aluminum sample. The right part of the sample is anodized.

6.3.3.5 Anodized aluminum

The final measurement on a coated system that will be discussed here is the case of anodized aluminum. In figure 6.24, a potential scan over partly anodized aluminum is shown. The figure shows that the potential above anodized aluminum is clearly lower (about 0.6V) than above the aluminum with an ordinary Al_2O_3 barrier layer. This is probably due to differences of the surface dipoles between the anodized and barrier layers together with possible differences in the adhesion voltages.

The aluminum sample was anodized in a diluted sulfuric acid in water, with 180g of H_2SO_4 per liter. The current density was $2\text{A}/\text{dm}^2$. In this way, an anodized layer of about $2\mu\text{m}$ is obtained in about 4 minutes [Spoelstra 1999].

6.4 Summary and conclusions

The original reason for studying the SKP was its possible use for studying localized corrosion. However as discussed in chapter 5, the value of the SKP for corrosion research is limited due to ambiguities in the interpretation of corrosion potential scans. However, the discussions in chapter 5 also suggested the application of the SKP for measuring the local adhesion differences of insulating coatings on metals, which is interesting for our study on the influence of the DWI process on PET coated ECCS. This was a new motivation for studying the SKP. We have studied the SKP itself, as well as its application. The study on the apparatus may be summarized as follows:

- The SKP does measure surface potentials. This was verified by a carbon sheet over which a known voltage was put and by an aluminum sample with a copper line on it. The latter sample was also used for resolution measurements;
- Drifting of the measured potential is probably due to temperature variations in the backward electronic circuit, which causes the CPDs in the backward circuit to vary. Controlling the temperature of electronics may therefore improve the stability;
- Slope distance control is a better way of capacitive distance control than are second harmonics and high U distance control, since it does not need high voltages and is not influenced by potential variations on the substrate. The nonlinearity of the probe current was given in terms of the capacity between the probe and the substrate, which was used for describing second harmonics distance control;
- Several laser methods were also considered for distance control. An attractive option is to use a combined probe-wave guide by sputtering a metal layer on an optical fiber. When a lens is etched on the tip, one may use the autofocus system of a CD player to control the distance between the probe and the substrate;
- Measurements on conducting samples can well be described in terms of mutual capacities. These will not be known normally, but it may payoff to determine them by finite element analysis since the mutual capacities can be used for improving potential scans by inversion techniques and for more accurate adjustment of the height during distance control. Finite element analysis can also be helpful for a better study of in the influence of coatings, which can only be done in one dimension by analytical methods.

In addition, a number of measurements on metal-electrolyte systems and on metals with different insulating layers were performed. This led to the following results.

- The corrosion potential alone is no good measure for the corrosion activity, as follows from measurements on iron in diluted sulfuric acid of different pH values. Measurements on filiform corrosion of other workers further show that potential differences are no good measure for the local corrosion activity. These observations are in line with the discussion in chapter 5. In addition, the electric potential of a surface is influenced by many other factors, which often have a much larger contribution to the surface potential than corrosion phenomena;
- Studying the local adhesion of PET on ECCS, one of our main goals of applying the SKP is seriously complicated by the extremely poor electric conductivity of PET. This makes it very difficult to dissipate static charges. Therefore the presented potential scans on PET coated ECCS may be criticized;
- Another way of measuring local adhesion is to homogeneously charge the coating by the corona method and then follow the local charge decay by the SKP at an elevated temperature. The advantage of this technique is that the associated surface potentials are so high that they by far dominate the other contributions. Some preliminary measurements were done, but the method needs further study;
- Studying other coated systems would be interesting. However potential scans on coated metals may often lead to surprising results, as for instance illustrated by our measurements on metals covered by glass slides or Scotch tape[®]. This once more shows the difficulty to interpret surface potential scans.

It follows that the versatility of the SKP is in fact also its handicap in case one wants to relate measured surface potential differences to certain specific phenomena. Nevertheless, the surface potential shows different details than other microscopic techniques and in that sense, the SKP may usefully contribute to studies of conducting surfaces as well as insulating surfaces.



Chapter 7

Inversion of dielectric spectra for obtaining time constant distributions

7.1 Introduction

The electric susceptibility χ is a material property and a measure for the electric polarizability of a material. In System International units (SI-units), χ is defined by the relation between the electric polarization P of a medium and the electric field E acting on it (cf. [Lorrain 1962]):

$$P = \epsilon_0 \chi E$$

Here the constant $\epsilon_0 = 8.9 \times 10^{-12}$ F/m is the permittivity of vacuum. We will only consider isotropic media, for which χ is a scalar. Debye postulated that the electric polarization of a material is due to the orientation of permanent electric dipoles inside the material [Debye 1912]. He also studied the frequency dependence of the susceptibility and came with the following result:

$$\chi(\omega) = \frac{\chi(0)}{1 + i\omega\tau}$$

Here ω is the angular frequency of the applied electric field, and τ is the time constant of the dipole relaxation. The susceptibility at zero frequency is real, and at higher frequencies it is a complex quantity, relating the complex polarization to the complex electric field in the usual treatment of linear time-invariant systems. The Debye susceptibility can be separated in real and imaginary parts:

$$\chi(\omega) = \chi(0) \left[\frac{1}{1 + (\omega\tau)^2} - \frac{i\omega\tau}{1 + (\omega\tau)^2} \right] = \chi'(\omega) - i\chi''(\omega)$$

When ω increases from 0 to ∞ , the real part $\chi'(\omega)$ decreases monotonously from $\chi(0)$ to 0, while the imaginary part $\chi''(\omega)$ has a maximum at $\omega = 1/\tau$ and vanishes for $\omega \rightarrow 0$ and $\omega \rightarrow \infty$.

∞ . The quantities $\chi'(\omega)$ and $\chi''(\omega)$ are often referred to as the storage and loss functions. They are plotted in figure 7.1 for a time constant $\tau = 0.01$ s. The chosen frequency interval from 10^{-2} Hz to 10^6 Hz is a common frequency spectrum.

For materials with several types of permanent dipoles, the polarizations add up and with that, the susceptibilities. Polymers are a good example of materials with several types of permanent dipoles. Typically, the polarization of a polymer is caused by the orientation of the net dipole of the main chain (called the α -relaxation) and the dipole moment of the side group, if present (called the β -relaxation). One might expect that the susceptibility of such a polymer could be described by the summation of two Debye relaxations with different amplitudes $\chi_\alpha(0)$ and $\chi_\beta(0)$ and with different time constants τ_α and τ_β :

$$\chi(\omega) = \frac{\chi_\alpha(0)}{1 + i\omega\tau_\alpha} + \frac{\chi_\beta(0)}{1 + i\omega\tau_\beta}$$

For actual polymers however, this is usually no accurate description. This is explained by the amorphous, random structure of polymers, which causes the environment of a single type of dipole to vary considerably within the polymer. Thus, the time constant of each dipole relaxation also varies considerably and the resulting susceptibility is a large summation of Debye relaxations with different time constants. For a macroscopic specimen, this summation can be represented by an integral:

$$\chi(\omega) = \int_{\tau=0}^{\infty} \frac{f(\tau)d\tau}{1 + i\omega\tau} \quad (7.1)$$

where $f(\tau)d\tau$ is the amplitude of the relaxation with time constant τ . The function $f(\tau)$ is hence the distribution of relaxation times (cf. [Böttcher, 1978]). The relaxation time distribution function is an interesting quantity to study because it contains information on the environment of the dipoles inside the medium. For instance, one can study how $f(\tau)$ changes by a heat treatment, aging or by water uptake.

Determination of the time constant distribution $f(\tau)$ requires the inversion of integral equation (7.1). In this chapter, we focus on this problem and study the following three inversion methods:

- Least square fitting;
- Inversion by differential operators;
- Deconvolution by Fourier transforms.

The main ideas of these methods are explained in section 7.2 as well as the implementation into operational algorithms. In section 7.3, the general limitations of the inversion algorithms are discussed and finally in section 7.4, the performances of the algorithms are compared by the inversion of some typical susceptibility spectra.

7.2 Three inversion methods

Measurements of $\chi(\omega)$ are normally done on a logarithmic ω -scale. For instance $\chi(\omega)$ may be measured for $\omega = 10^{1.0}, 10^{1.5}, 10^{2.0}, 10^{2.5}$ Hz and so on. The ω values are then a constant factor apart or in other words, a constant step $\Delta \log(\omega)$. Also the relaxation times are generally considered on a logarithmic scale. Therefore it is natural to write $\chi(\omega)$ as an integral over $\ln(\tau)$:

$$\chi(\omega) = \int_{\ln \tau = -\infty}^{\infty} \frac{g(\ln \tau) d \ln \tau}{1 + i\omega\tau} \tag{7.2}$$

where $g = \tau f$ is the logarithmic time constant distribution function. Separation in real and imaginary parts leads to:

$$\chi'(\omega) = \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) \frac{1}{1 + (\omega\tau)^2} d \ln \tau \equiv \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) h'(\omega\tau) d \ln \tau \tag{7.3a}$$

$$\chi''(\omega) = \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) \frac{\omega\tau}{1 + (\omega\tau)^2} d \ln \tau \equiv \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) h''(\omega\tau) d \ln \tau \tag{7.3b}$$

where the storage and loss kernels $h'(\omega\tau)$ and $h''(\omega\tau)$ are defined. It follows that $g(\ln \tau)$ can be determined from either the real or the imaginary part of $\chi(\omega)$. We will study both options and now proceed by a discussion on the main ideas of the three methods mentioned in the introduction.

7.2.1 Least square fitting

In practice, always a discrete number of N susceptibility values $\chi_i = \chi(\omega_i)$ will be available for the inversion. The idea of this algorithm is now to fit M values of $g_j = g(\tau_j)$ to the data points by considering the discrete versions of (7.3a) and (7.3b):

$$\chi_i' = \sum_{j=1}^M h_{ij}' g_j \Delta \ln \tau$$

$$\chi_i'' = \sum_{j=1}^M h_{ij}'' g_j \Delta \ln \tau$$

The maximum number of g_j values that can be fitted is N , in which case h' and h'' are square matrices and the solution for g_j is unique (provided that h' or h'' are regular). For reasons that will be discussed in section 7.2.3, it is however generally wiser to choose M considerably smaller than N . The representation of the χ_i data by the set of g_j values is then generally not exact and denoting the "errors" by e_i' and e_i'' respectively, we have:

$$\chi_i' = \sum_{j=1}^M h_{ij}' g_j \Delta \ln \tau + e_i'$$

$$\chi_i'' = \sum_{j=1}^M h_{ij}'' g_j \Delta \ln \tau + e_i''$$

The optimal g_j values can now be defined as those values minimizing the length of the error vectors. This leads to the least square estimates:

$$\hat{g}' \Delta \ln \tau = [(h')^T h']^{-1} [h']^T \chi' \quad (7.4a)$$

$$\hat{g}'' \Delta \ln \tau = [(h'')^T h'']^{-1} [h'']^T \chi'' \quad (7.4b)$$

Weeze and Honerkamp for instance applied this inversion method [Weeze 1993, Honerkamp 1993]. The main parameters are the number of time constants M for which g_j values are fitted and the spacing $\Delta \ln \tau$ between the time constants.

7.2.2 Inversion by differential operators

After some substitutions, the integral equation (7.2) can be put into the following form:

$$f(s) = \int_{t=0}^{\infty} \frac{1}{s+t} \phi(t) dt$$

This is the Stieltjes transform, which can be inverted by the following differential operator [Widder 1946]:

$$\phi(x) = \lim_{n \rightarrow \infty} (-1)^{n-1} \frac{1}{\Gamma(n+1)\Gamma(n-1)} \frac{d^n}{dx^n} \left[x^{2n-1} \frac{d^{n-1} f(x)}{dx^{n-1}} \right]$$

Love and Byrne gave some variations on this inversion formula [Love 1980]. For finite values of n , Widder's operator is an approximation, on which another class of inversion algorithms for (7.3a) and (7.3b) is based. Especially in the field of viscoelastic and dielectric spectroscopy, differential operator algorithms are popular [Schwarzl 1953, Ferry 1968, Böttcher 1978, Tschoegl 1989, Friedrich 1991, Emri 1993]. Schwarzl and Staverman had an interesting and different approach to the inversion of (7.3a) and (7.3b), which they only worked out up to a fourth order differential operator. Their ideas were carried on by Ferry and Tschoegl and finally put into a closed form by Friedrich [Friedrich 1991]:

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} (-1)^{(n-1)/2} \frac{2}{\Gamma\left[\frac{1}{2}(n+1)\right]^2} \omega^{n+1} \frac{d^n}{d(\omega^2)^n} \left[\omega^{n-1} \chi'(\omega) \right]$$

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} (-1)^{n/2} \frac{2}{\Gamma\left[\frac{1}{2}(n+1)\right]^2} \omega^{n+1} \frac{d^n}{d(\omega^2)^n} \left[\omega^{n-1} \chi''(\omega) \right]$$

where $1/\omega$ is the τ value. These expressions are more easily implemented than Widder's original operator, because a higher order derivative has to be evaluated only once. In this work, we will however use the following expressions instead:

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} \prod_{n=\text{even}} \left(1 - \frac{1}{n^2} \frac{d^2}{d(\ln \omega)^2} \right) \left[-\frac{d\chi'(\omega)}{d(\ln \omega)} \right] \quad (7.5a)$$

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} \prod_{n=\text{odd}} \left(1 - \frac{1}{n^2} \frac{d^2}{d(\ln \omega)^2} \right) \left[\frac{2}{\pi} \chi''(\omega) \right] \quad (7.5b)$$

which are even more suited for implementation, due to the fact that the differential operators are factorized in second order differentiations. Hence, the approximation of order $(n+2)$ can be obtained from the approximation of order n by a second order differentiation. For Friedrich's operators, this is not possible and one has to start all over again for each higher order approximation. Since in practice one does not know how large n should be for convergence, equations (7.5a) and (7.5b) are therefore more practical. In addition, the inversion operators of (7.5a) and (7.5b) are based on logarithmic differentiations. This is also convenient since in most cases, the ω array will be logarithmically spaced. A proof of (7.5a) and (7.5b) is given in appendix A.

Implementation to computer programs

The main issue for this inversion method is how to calculate the first and second order derivatives from the data arrays. For an array of N points, we did this as follows:

- Choose a number M defining the width of a differential interval. This leads to $(N-M+1)$ intervals from $1..M$, $2..(M+1)$, and so on up to $(N-M+1..N)$;
- For the first derivative, fit a straight line through the M points of the intervals. The slope of this line is now the first derivative on the midpoint of the interval. Since ω is logarithmically spaced, the midpoint is at $\omega = \sqrt{\omega_1 \omega_M}$. A similar procedure is followed for the second derivative, in which case a parabola is fitted.

After each iteration (or second order differentiation step), the frequency values are replaced by the midpoint values for which the derivative is calculated. After the last iteration (of course only a finite, specified number of iterations is carried out), the several times differentiated data array approximates the time constant distribution and the inverses of the final ω values are taken as the τ values. Since the data array loses $M-1$ points per differentiation step, the resulting time constant distribution g_τ will contain considerably less points than the original data arrays χ_i' and χ_i'' . One can try to avoid the loss of data points on the boundaries of the spectrum by differentiations with fewer points or by asymmetric differentiations, but this will lead to unreliable results for the highest and lowest τ values and is not recommended.

Instead of polynomial fitting, several of the authors mentioned above (cf. [Ferry 1968, Tschoegl 1989]) used closed expressions for M -points differences for approximating the derivatives. This however comes down to the same thing as polynomial fitting through M points and is more work because all these closed expressions have to be specified. Especially

when a computer carries out the algorithm, it makes no sense to do this and it is easier to calculate the derivatives by polynomial fitting.

It is sometimes mentioned (cf. [Tschoegl, 1989]) that too many data points per frequency decade can have a negative influence on the final result. The explanation of this phenomenon is that when the ω values are very densely spaced, they add no information, while at the same time the τ values become too densely spaced. As a result we get the effect of fitting too many parameters to the data. To overcome this problem, we introduced an additional "decimation parameter" that removes a certain fraction of points from the ω array between successive differentiation steps.

7.2.3 Deconvolution by Fourier transforms

Inversion problems such as equation (7.2) can also be considered in the framework of Fourier transforms. This leads to inversion algorithms (cf. [Meinke 1959, McWhirther 1978, Link 1991]) and also to a powerful method for studying the uniqueness of the solution (cf. [McWhirther 1978, Parker 1994]). The latter issue will be discussed in section 7.3.

The first step of this inversion method is to write (7.2) in the form of a convolution integral. This is done by the substitutions $p = \ln(\omega)$ and $q = -\ln(\tau)$:

$$\chi(p) = \int_{q=-\infty}^{\infty} \frac{g(q) dq}{1 + i e^{p-q}}$$

with the corresponding loss and storage functions:

$$\chi'(p) = \int_{q=-\infty}^{\infty} \frac{1}{1 + e^{2(p-q)}} g(q) dq = \int_{q=-\infty}^{\infty} h'(p-q) g(q) dq \quad (7.6a)$$

$$\chi''(p) = \int_{q=-\infty}^{\infty} \frac{e^{p-q}}{1 + e^{2(p-q)}} g(q) dq = \int_{q=-\infty}^{\infty} h''(p-q) g(q) dq \quad (7.6b)$$

These are both convolution integrals. The idea is now to take the Fourier transforms of the equations by which the convolution integrals are changed into simple products. For (7.6b) this leads to the following:

$$X''(k) = H''(k)G(k)$$

where $X''(k)$, $H''(k)$ and $G(k)$ are the Fourier transforms of $\chi''(p)$, $h''(p)$ and $g''(p)$ respectively. Here, the Fourier transform of $h''(p)$ is defined by:

$$H''(k) = F_k \{h''(p)\} = \int_{p=-\infty}^{\infty} h''(p) e^{-ikp} dp$$

and similar expressions hold for $X''(k)$ and $G(k)$. In the k -space (or Fourier-space), the inversion is simple:

$$G(k) = X''(k)/H''(k)$$

and the time constant distribution finally follows by an inverse Fourier transform:

$$g(q) = F_q^{-1}\{G(k)\} = \frac{1}{2\pi} \int_{k=-\infty}^{\infty} G(k)e^{ikq} dk$$

completing the inversion. An analytical expression for $H''(k)$ can be given:

$$H''(k) = F_k \left\{ \frac{1}{2 \cosh(p)} \right\} = \frac{\pi/2}{\cosh(\pi k/2)}$$

which makes the whole procedure even more convenient. One would like to follow a similar procedure for the storage peak $\chi'(p)$, but this is not so obvious. The main reason is the fact that $\chi'(p)$ vanishes for $p \rightarrow +\infty$ while $\chi'(p)$ is nonzero for $p \rightarrow -\infty$. This mismatch makes that the Fourier transform of $\chi'(p)$ cannot be determined accurately from actual data, as will be discussed below. There are several ways out, such as multiplying (7.6a) by $\exp(\beta p)$ with $0 < \beta < 1$ [McWhirther 1978, Link 1991]. Here, we choose a different option, which is to consider the derivative $d\chi'/dp$:

$$\frac{d\chi'}{dp} = - \int_{q=-\infty}^{\infty} \frac{1}{2 \cosh^2[2(p-q)]} g(q) dq$$

Now since $d\chi'/dp$ vanishes for both $p \rightarrow +\infty$ and $p \rightarrow -\infty$, there is no problem with the calculation of its Fourier transform. The Fourier transform of the kernel of the latter convolution integral can again be given in a closed form:

$$F_k \left\{ \frac{1}{2 \cosh^2(2p)} \right\} = \frac{\pi k/2}{\sinh(\pi k/2)}$$

Here for $k = 0$, the limiting value 1 should be taken. Summarizing, we have the following inversion formulas:

$$g(q) = F_q^{-1} \left\{ \frac{\sinh(\pi k/2)}{\pi k/2} F_k \left\{ -\frac{d\chi'}{dp} \right\} \right\} \tag{7.7a}$$

$$g(q) = F_q^{-1} \left\{ \frac{\cosh(\pi k/2)}{\pi/2} F_k \{ \chi''(q) \} \right\} \tag{7.7b}$$

Before discussing the implementation of this method to an operational algorithm, we like to mention an interesting consequence of the latter equations. Since both are equal to $g(q)$, their right hand sides are equal and one can write the following:

$$F_k \{ \chi''(p) \} = \frac{\tanh(\pi k / 2)}{k} F_k \left\{ -\frac{d\chi'}{dp} \right\} \quad (7.8)$$

This provides a way to calculate $\chi''(p)$ from $\chi'(p)$ and vice versa, apart from a constant offset in $\chi'(p)$ in the latter case. This constant must be determined from an additional constraint such as that $\chi'(p)$ vanishes for $p \rightarrow \infty$. Equation (7.8) is equivalent to the Kramers-Kronig transforms that are also used for the determination of $\chi''(p)$ from $\chi'(p)$ and vice versa. This is done for experiments by which only $\chi''(p)$ or $\chi'(p)$ can be measured or for the validation of dielectric or electric impedance spectra (cf. [Böttcher 1978, Jonscher 1983, Macdonald 1987]).

Implementation to computer programs

For actual susceptibility data, a “discrete Fourier transform” (DFT) must approximate the continuous Fourier transform. The “fast Fourier transform” (FFT) is the most popular algorithm for obtaining the DFT. The following pair of equations defines the DFT and its inversion:

$$X_n = \sum_{m=0}^{N-1} x_m e^{-im\frac{2\pi}{N}} \quad (7.9a)$$

$$x_m = \frac{1}{N} \sum_{n=0}^{N-1} X_n e^{im\frac{2\pi}{N}} \quad (7.9b)$$

Here X_n is the DFT, x_m the signal and N the number of data points. As $N \rightarrow \infty$, these equations must gradually turn into the continuous Fourier transform and its inversion:

$$X(k) = \int_{-\infty}^{\infty} x(p) e^{-ikp} dp$$

$$x(p) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(k) e^{ikp} dk$$

This is achieved by the following list of substitutions:

$$p = m\Delta p$$

$$k = n\Delta k$$

$$x_m = x(m\Delta p) = x(p)$$

$$X_n = X(n\Delta k) = X(k)$$

$$\Delta p \Delta k = 2\pi / N$$

The lower equation is sometimes called the “uncertainty relation”. This is a relation between the step sizes in the k -space and the p -space.

Since the data array x_m consists of N points, X_n will consist of N points too. But since x_m is real and X_n is complex with real and imaginary parts $\text{Re}\{X_n\}$ and $\text{Im}\{X_n\}$, it may seem

as if X_n contains two times more information than x_m . However it follows from (7.9a) and the fact that x_m is real, that $X_{N-n} = X_n^*$, hence that $\text{Re}\{X_{N-n}\} = \text{Re}\{X_n^*\}$ and $\text{Im}\{X_{N-n}\} = -\text{Im}\{X_n^*\}$. By consequence, $\text{Im}\{X_0\} = 0$ and for even N , $\text{Im}\{X_{N/2}\} = 0$ too. It follows that all information of the DFT is contained by:

- $\text{Re}\{X_0..X_{N/2}\}$ and $\text{Im}\{X_1..X_{N/2-1}\}$ for even N ;
- $\text{Re}\{X_0..X_{(N-1)/2}\}$ and $\text{Im}\{X_1..X_{(N-1)/2}\}$ for odd N .

In both cases these are exactly N independent real numbers, hence X_n contains as much information as x_m .

Now we come back to the main reason why it is better to work with $d\chi'/dp$ than with $\chi'(p)$. From (7.9a), it follows that $x_m = x_{m+N}$, hence the actual signal x_m from which the DFT is calculated must be regarded as a periodic signal. But this means that the data points x_0 and x_{N-1} for the lowest and highest frequencies are put "head to tail", which comes down to a huge discontinuity in case of the $\chi'(p)$ data. This discontinuity gives a major, but fully artificial contribution to the calculated DFT and finally leads to completely wrong results. Therefore it is better to work with $d\chi'/dp$, which quantity vanishes for both $p \rightarrow -\infty$ and $p \rightarrow +\infty$, similar to $\chi''(p)$.

There are other reasons why it may not be possible to obtain the Fourier transform with a very high accuracy. These all have to do with some kind of "incompleteness" of the data. The following examples of incompleteness are of concern:

- The frequency spectrum is too narrow to contain the whole $\chi(\omega)$. Then $d\chi'/d(\ln\omega)$ and $\chi''(\omega)$ are significantly nonzero beyond the boundaries of the spectrum;
- The frequency spacing of the spectrum may be too wide to describe the finest details in $\chi'(\omega)$ and $\chi''(\omega)$. This effect is called "undersampling";
- The noise level may be too high and hides the finer details in $\chi'(\omega)$ and $\chi''(\omega)$.

In general, the time constant distribution $g(q)$ will be a moderately smooth function and therefore its Fourier transform $G(k)$ will be the largest for the smaller k values. In other words, the modulus $|G(k)|$ will generally decrease with $|k|$. By consequence, the perturbing effect due to incompleteness will generally be most severe for the higher k values. Therefore the direct application of (7.7a) or (7.7b) will generally lead to totally wrong results for $g(q)$, even for rather complete data. A convenient way out is to ignore $G(k)$ for the higher k values. This comes down to low pass filtering of $G(k)$, or the multiplication by a window $W(k)$ prior to the inverse Fourier transform:

$$g(q) = F_q^{-1}\{W(k)G(k)\}$$

Many different low-pass filters are used in practice (cf. [Priestly 1981]). We worked with the rectangular window:

$$W^R(k) = 1, \text{ for } |k| < k_{\max}$$

$$W^R(k) = 0, \text{ for } |k| > k_{\max}$$

and with the Tukey-Hanning window:

$$W^{TH}(k) = \frac{1}{2} [1 + \cos(\pi k / k_{\max})], \text{ for } |k| < k_{\max}$$

$$W^{TH}(k) = 0, \text{ for } |k| > k_{\max}$$

The value k_{\max} is the cutoff k value and is the most important filtering parameter. In practice, its value can be estimated from visual inspection of the unfiltered $G(k)$. In section 7.4, this procedure is demonstrated.

The Tukey-Hanning window is known for its smoothing abilities. This is illustrated by considering the Tukey-Hanning window for the largest possible $k_{\max} = \pi/\Delta p$. Then:

$$W^{TH}(n\Delta k) = W_n = \frac{1}{2} [1 + \cos(2\pi n / N)] = \sum_{m=0}^{N-1} w_m e^{-imn \frac{2\pi}{N}}$$

By expansion of the cosine in complex exponentials, it follows that the corresponding window $w_m = w(m\Delta p)$ in the p -space reads:

$$w_m = \frac{1}{4}, \text{ for } m = 1 \text{ and } m = -1$$

$$w_m = \frac{1}{2}, \text{ for } m = 0$$

and $w_m = 0$ for all other values. Now since multiplication in the k -space corresponds to convolution in the p -space, application of the Tukey-Hanning window comes down to the following operation:

$$g_j \rightarrow \frac{1}{4} g_{j-1} + \frac{1}{2} g_j + \frac{1}{4} g_{j+1}$$

Hence, g_j is turned into a weighed average with the neighboring points, which is clearly a smoothing procedure. For values of k_{\max} lower than $\pi/\Delta p$, the smoothing is carried out over more than three points.

7.3 Limitations of the inversion algorithms

We will now discuss whether the inversion problems (7.3a) and (7.3b) for determining $g(\ln r)$ from either $\chi'(\omega)$ or $\chi''(\omega)$ have a unique solution. This is very important for the success of the inversion algorithms. Both inversion problems require the inversion of an integral equation of the following form:

$$\phi(x) = \int_{x'-a}^b h(x, x') s(x') dx' \quad (7.10)$$

This integral equation is a Fredholm integral equation of the first kind. The problem is also known as the "inverse source problem", where the goal is to determine the unknown source

distribution $s(x)$ from a detected signal $\phi(x)$. The kernel $h(x,x')$ then describes the response of a detector at position x to a source at position x' . A large variety of experiments can be put into this form (cf. [Parker 1994]).

It was already pointed out by Fredholm that equation (7.10) either has 0, 1 or infinitely many solutions for $s(x)$. The case of infinitely many solutions occurs solely when there are "non-radiating source distributions" $s_{NR}(x)$, having the property that:

$$\int_{x'=a}^b h(x, x')s_{NR}(x')dx' = 0$$

The detector response to a non-radiating source distribution is therefore zero and one could say that the detector does not "see" non-radiating sources. Now when there are non-radiating sources $s_{NR}(x)$ unequal to zero, a solution $s(x)$ of equation (7.10) cannot be unique since then, any source distribution $s(x) + \alpha s_{NR}(x)$ will also do for any α . Then, there are infinitely many solutions.

One way to examine whether a particular inverse source problem has nontrivial, non-radiating sources, is to write the Fredholm integral of the first kind into the form of a convolution integral (cf. [McWhirther 1978, Parker 1994]):

$$\phi(x) = \int_{x'=-\infty}^{\infty} h(x - x')s(x')dx'$$

For that purpose, a proper substitution of variables may be necessary. In case of a product kernel $h(xx')$, this is achieved by the substitution $x = \exp(p)$ and $x' = \exp(-q)$ as demonstrated in section 7.2.3. Then by a Fourier transform, the integral equation turns into a product:

$$F(k) = H(k)S(k)$$

where F , H , and S are the Fourier transforms of ϕ , h and s . Now suppose that $s_{NR}(x)$ is a non-radiating source. Then $\phi(x) = 0$ for all x . Then $F(k) = 0$ for all k and hence, the product $H(k)S_{NR}(k) = 0$ for all k . Now in case $H(k)$ is nonzero for all k , it follows that $S_{NR}(k) = 0$ for all k and by consequence that $s_{NR}(x) = 0$ for all x . Hence the condition that $H(k)$ is nonzero for all k is sufficient to guarantee the absence of nontrivial, non-radiating sources.

Following these considerations, it is rapidly shown that the inversion problems (7.3a) and (7.3b) both have a unique solution for $g(\ln t)$, since the transformation into convolution integrals was already done in section 7.2.3. This led to:

$$F_k \left\{ \frac{d\chi'(p)}{dp} \right\} = \frac{\pi k / 2}{\sinh(\pi k / 2)} G(k) \equiv H'(k)G(k)$$

$$F_k \{ \chi''(q) \} = \frac{\pi / 2}{\cosh(\pi k / 2)} G(k) \equiv H''(k)G(k)$$

Now since both $H'(k)$ and $H''(k)$ are nonzero functions of k , there are no nontrivial, non-radiating sources and the solution of both inversion problems (7.3a) and (7.3b) is unique. Hence one may expect that the inversion may be carried out rather successfully by either of

the methods of section 7.2.1, 7.2.2 or 7.2.3, provided that the susceptibility data are sufficiently complete. Note however that for the larger k values, both $H'(k)$ nor $H''(k)$ decrease rapidly. This means that the inversion will be particularly sensitive to the contamination of rapidly oscillating contributions, which will have to be suppressed by some kind of smoothing technique.

7.4 Comparison of the performance of the inversion algorithms

For testing and comparing the algorithms described in section 7.2, we used artificial susceptibility data, based on the "Cole-Cole susceptibility" [Cole 1941]:

$$\chi(\omega) = \frac{\chi(0)}{1 + (i\omega\tau_0)^\nu} \quad (7.11)$$

Here ν is a parameter between 0 and 1. For $\nu = 1$, the Cole-Cole susceptibility is identical to the Debye susceptibility. In figure 7.2, the real and imaginary parts of the Cole-Cole susceptibility are plotted for three different values of ν . The Cole-Cole susceptibility is used as a heuristic description of actual susceptibility data and therefore provides realistic data for testing. Another advantage of using Cole-Cole data is that an exact formula for the time constant distribution exists [Böttcher 1978]:

$$g(\ln \tau) = \frac{\chi(0)}{2\pi} \left[\frac{\sin(\nu\pi)}{\cosh[\nu \ln(\tau_0 / \tau)] + \cos(\nu\pi)} \right] \quad (7.12)$$

Hence we can compare the results of the inversion algorithms to the exact outcome. Equation (7.12) is valid for $0 < \nu < 1$. For $\nu \rightarrow 1$, the Dirac delta function is obtained as the susceptibility turns into the Debye susceptibility. In figure 7.3, the time constant distribution functions are plotted for the susceptibility spectra of figure 7.2.

7.4.1 Cole-Cole data without noise

The susceptibilities of figure 7.2 are our first testing objects. We took 201 frequency points equidistantially spaced over 8 decades. (For experimental data and especially for the lower frequencies for which the measurements are most time consuming, one will often take fewer points. This may also lead to good inversion results as long as the data are sufficiently complete.) The results of the least squares algorithm are shown in figures 7.4, 7.5 and 7.6. The solid lines, which are the analytical results, are not smooth because the function values were only calculated for the fitted time constants. As we see, the inversion is almost perfect for all ν values and for both the storage and loss data. Only for the smallest time constants small errors are visible. The number of τ values was determined by trial and error. Note that considerably fewer τ values could be fitted than the number of frequency points.

For the differentiation algorithm, the results are given in figures 7.7, 7.8 and 7.9. We see that the results are about perfect for $\nu = 0.4$ and $\nu = 0.6$, but that the algorithm has

problems with the sharpness of the peak for $\nu = 0.8$. The convergence of the algorithm for the case $\nu = 0.6$ is illustrated by figure 7.10. In this figure, the results are plotted after 0 to 6 iterations (which comes down to 6 second order differentiation steps). Only the convergence of the loss data is shown. After 6 iteration steps, there is hardly any further improvement, which is also illustrated by figure 7.11. Here the normalized kernels $\hat{h}_{nn}(\omega\tau)$ are plotted as calculated by the algorithm for $n = 0, 2, 4, 6, 8, 10, 12$. We see that the kernel actually does not become sharp enough and does not approach the Dirac delta function. Apparently, this makes it impossible to do the inversion for the case $\nu = 0.8$ very well. Since the minimum number of three points was used to calculate the second order derivatives, one may conclude that the frequencies were too widely spaced for this case.

Finally, the Fourier deconvolution algorithm seems to handle all three cases rather well, as shown in figures 7.12, 7.13 and 7.14. Only for the smallest and largest time constants, small errors are visible. Note that the results obtained from the storage data are somewhat better than the results of the loss data. Figure 7.15 illustrates the need of filtering, even for analytical data. In this figure, the solid line shows the Fourier transform of the exact time constant distribution, which falls off to zero with increasing $|k|$. The experimental results are however very incorrect for the larger k values, which must somehow be due to the finite width and spacing of the frequency spectrum. Based on this figure, the Fourier transforms were cut off at $k_{\max} = 3.0$ for the loss data and at $k_{\max} = 4.5$ for the storage data.

7.4.2 Spectra with noise

To study the influence of noise on the inversion, normally distributed white noise was added to the susceptibility data. The influence of the noise is illustrated by figures 7.16 and 7.17. These figures were calculated for a Cole-Cole susceptibility with $\nu = 0.6$, similar to figure 7.15, apart from the fact that noise was added to the susceptibility data with standard deviations 0.001 and 0.01 respectively. The loss of information due to the noise is clearly visible in the Fourier transforms, which have to be cut off at significantly lower k values.

In order to make a somewhat harder test for the inversion algorithms, a test spectrum with two different Cole-Cole relaxations was considered:

$$\chi(\omega) = \frac{0.5}{1 + (i\omega/3)^{0.5}} + \frac{1}{1 + (i\omega/3000)^{0.75}}$$

According to this formula, $\chi(\omega)$ was calculated for 201 frequencies, equidistantly spaced over 8 decades. Normally distributed noise with a standard deviation of 0.01 was added to the data and the result is plotted in figure 7.18. This test spectrum is meant as a simulation of the susceptibility of a polymer, with a sharper α relaxation and a smoother β -relaxation. The question was now whether the three algorithms could find the hidden time constants at $1/3$ s and at $1/3000$ s as well as the global form of the time constant distribution function.

In figure 7.19, the result of the least square fitting is shown, which is rather good. The influence of noise is suppressed by simply fitting fewer time constants. In figure 7.20, the result for the differential operator is shown. The best result was obtained when 11 points were taken for the second order differentiations. No more than a single iteration was done for the storage data and two were done for the loss data. A decimation of 2 points was carried out to

relax the time constant intervals. The significance of the decimation is illustrated by figure 7.21. In this case the decimation was omitted, which clearly leads to a worse result.

The results from the Fourier deconvolution algorithm are shown in figures 7.22 and 7.23. The difference between the figures is that in the first case, a rectangular window was used and in the second case, a Tukey-Hanning window. The figures illustrate the smoothing abilities of the Tukey-Hanning window. However, the smoothing also flattens the sharper peak at 1/3000 s. The rectangular window gives a better representation of the sharper peak, but makes a mess of the rest of the spectrum.

7.5 Summary and discussion

We have discussed three methods for the calculation of time constant distributions from electric susceptibility spectra. These methods were:

- Least square fitting;
- Inversion by differential operators;
- Deconvolution by Fourier transforms.

All three methods were implemented in operational algorithms by using Matlab[®], and were tested on a number of artificial spectra for which the time constant distributions were known. Some new ideas were added to the methods. In particular, we used a new, factorized differential operator for the second method by which higher order approximations could be done more easily and we worked with the logarithmic derivative of the real part of the susceptibility for the third method. In addition, a new relation between the real and imaginary part of the susceptibility was found. This relation is equivalent to the Kramers-Kronig transforms and could be used whenever these equations apply.

The time constant distribution can be calculated from either the real or the imaginary part of the susceptibility, which in both cases comes down to the inversion of a Fredholm integral equation of the first kind. By Fourier analysis, it can be shown that both inversion problems have a unique solution. This means that the inversion can be carried out successfully, provided that the susceptibility spectrum is sufficiently complete. The latter was verified by the inversion of a number of wide susceptibility spectra without noise and with many frequencies and indeed all inversion methods gave rather good results.

The influence of noise was studied by adding normally distributed, white noise to the susceptibility spectra. All three algorithms include some kind of smoothing technique for the suppression of noise. The least squares algorithm does this by fitting fewer and more-widely spaced time constants, the differential operator algorithm by larger differentiation intervals and the Fourier deconvolution algorithm by low-pass filtering. The latter procedure is the most "mechanical" and requires basically the choice of a proper cutoff value by visual inspection of the unfiltered Fourier transform of the time constant distribution. The smoothing techniques of the other two algorithms are less transparent and for a great deal based on trial and error. This especially holds for the differential operator algorithm, for which also the number of differentiation steps must be chosen as well as a "decimation parameter" in order to arrive at a wider spacing of time constants. For the Fourier

deconvolution, the Tukey-Hanning window led to a smoother result than a rectangular window, but also to a worse representation of the sharper peaks in the time constant distribution.

For the inversion of analytical data, for which the frequency spectrum is arbitrarily wide and dense, the differential operator algorithm gave the best results although the results of the two other algorithms were also quite good. Hence for analytical susceptibility spectra, for which for instance no analytical expression for the time constant distribution is known, the differential operator algorithm is recommended.

However the least square or Fourier deconvolution algorithms, which do the inversion "at once" instead of by a number of iteration steps, are more efficient with the data and better for the inversion of measured susceptibility spectra. The disadvantage of the least square algorithm is then that the data processing requires some trial and error, which contrasts with the elegant low-pass filtering procedure of the Fourier deconvolution method. On the other hand, the least square algorithm can be extended more easily for additional constraints or "a priori" information such as the non-negativity of the time constant distribution function. Then the inversion problem is turned into an optimization problem, for which the parameter space can be confined and the squared error be extended by extra terms. In fact it seems to be that even without these extensions, the least squares algorithm is somewhat better than the Fourier deconvolution algorithm, but this should be verified by future reseach. Finally it would be interesting to apply the presented methods to other inversion problems of the same kind.

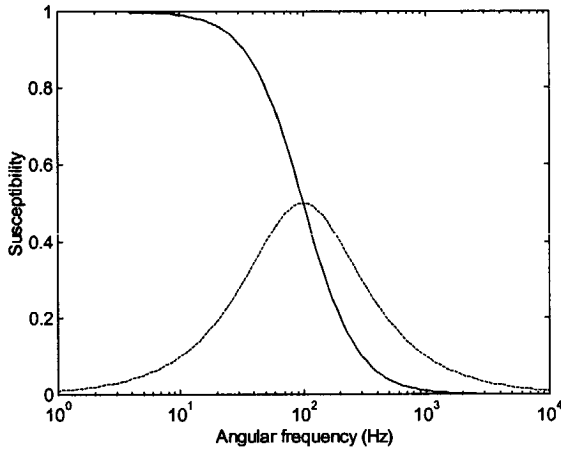


Figure 7.1. Debye relaxation. The solid line is the real part of the susceptibility and the dashed line is minus the imaginary part.

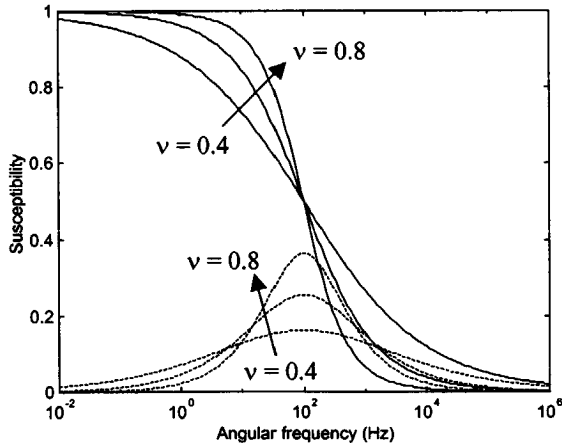


Figure 7.2. Cole-Cole susceptibility for $\nu = 0.4, 0.6, 0.8$ and for $\tau_0 = 0.01$ s. The solid line is the real part of the susceptibility, the dashed line is minus the imaginary part.

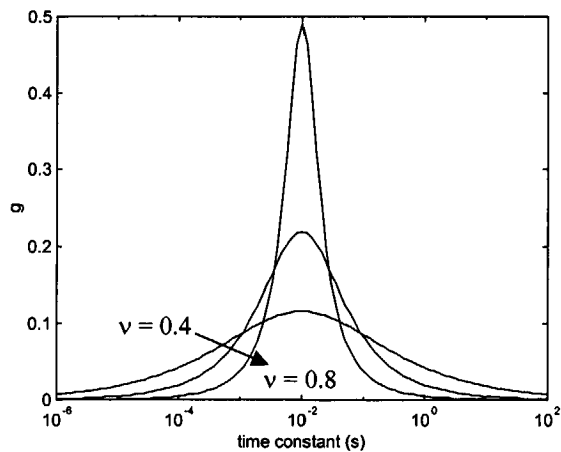


Figure 7.3. Cole-Cole time constant distribution for $\nu = 0.4, 0.6, 0.8$. For $\nu = 1$, the time constant distribution is a Dirac delta function.

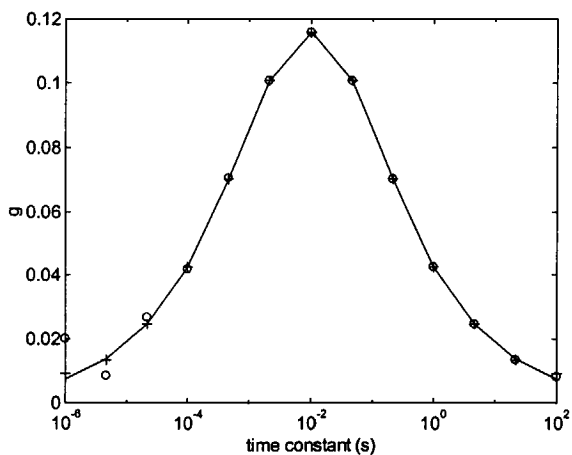


Figure 7.4. Least square fit for a Cole-Cole relaxation with $\nu = 0.4$. The solid line is the analytic result, the circles show the inversion from the storage data and the plusses give the inversion from the loss data.

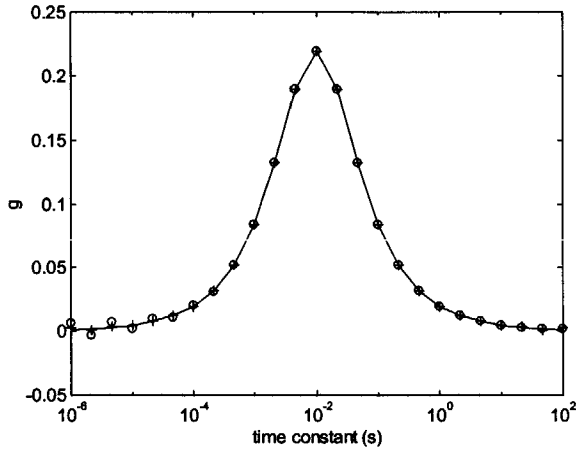


Figure 7.5. Least square fit for a Cole-Cole relaxation with $\nu = 0.6$. The solid line, the circles and the plusses have the same meaning as in the previous figure.

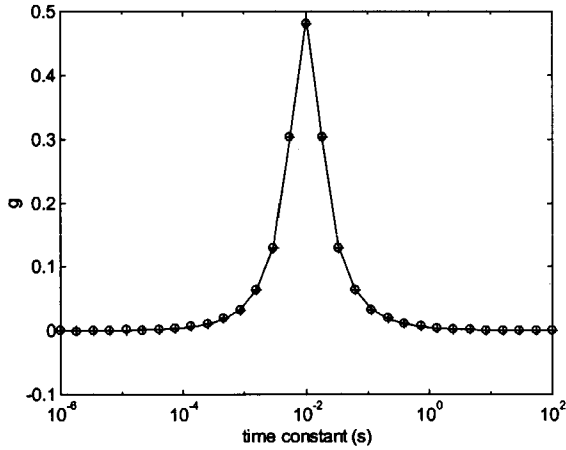


Figure 7.6. Least square fit for a Cole-Cole relaxation with $\nu = 0.8$. The solid line, the circles and the plusses have the same meaning as in the previous figure.

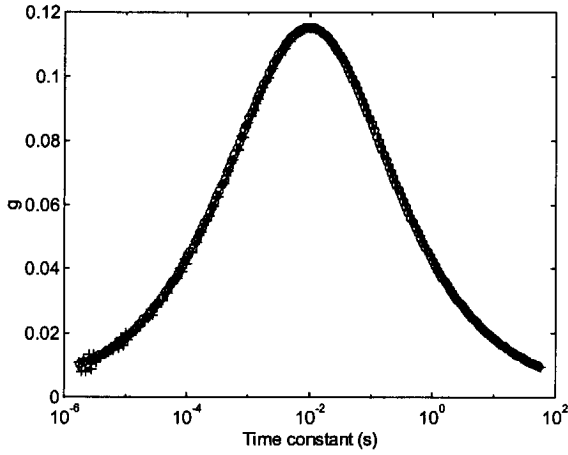


Figure 7.7. Differentiation result for a Cole-Cole relaxation with $\nu = 0.4$. The solid line, the circles and the plusses have the same meaning as in the previous figure.

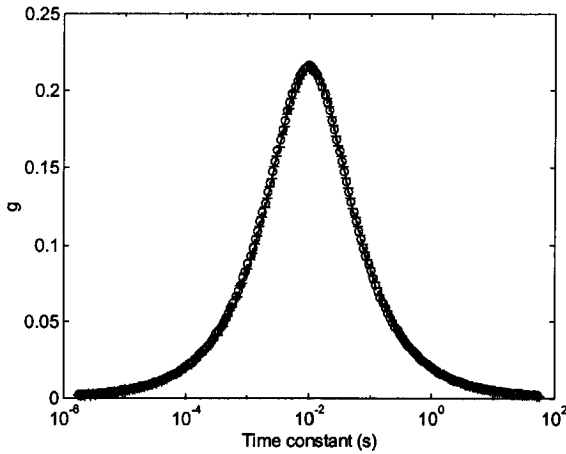


Figure 7.8. Differentiation result for a Cole-Cole relaxation with $\nu = 0.6$. The solid line, the circles and the plusses have the same meaning as in the previous figure.

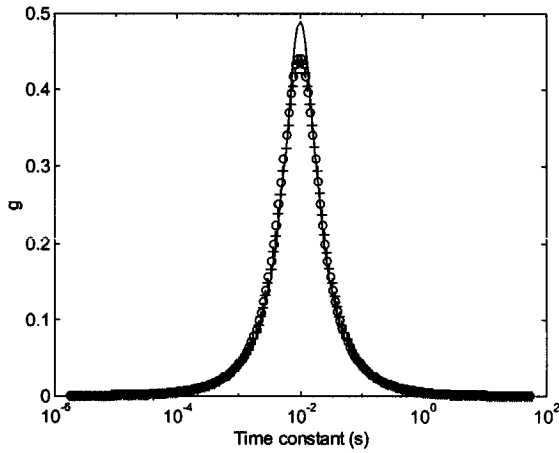


Figure 7.9. Differentiation result for a Cole-Cole relaxation with $\nu = 0.8$. The solid line, the circles and the plusses have the same meaning as in the previous figure.

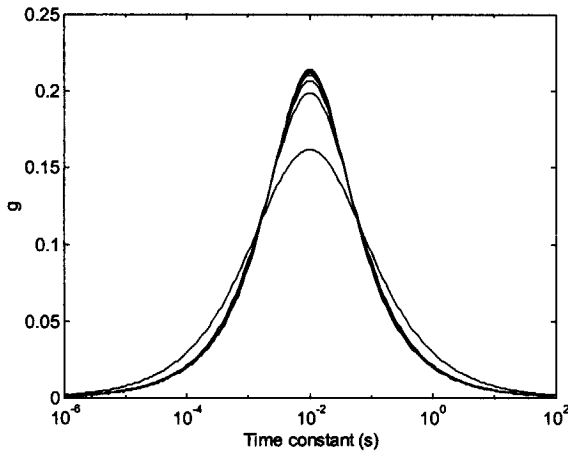


Figure 7.10. Convergence of the differentiation algorithm for a Cole-Cole relaxation with $\nu = 0.6$. The results after 0 to 6 iterations are plotted

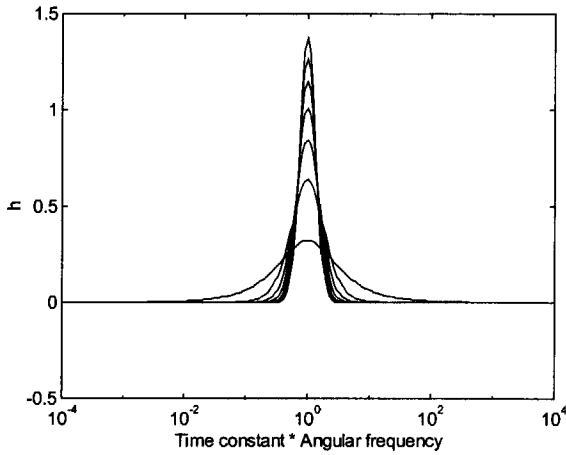


Figure 7.11. Normalized kernels, after 0 to 6 sharpening steps. Eventually the result should approach a Dirac delta function.

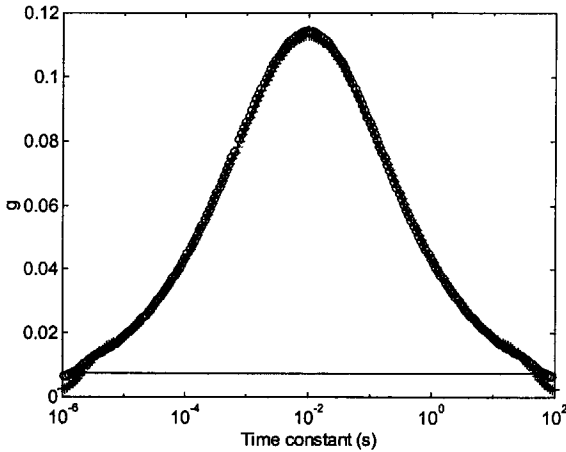


Figure 7.12. Deconvolution result for a Cole-Cole relaxation with $\nu = 0.4$. A second order filter is used with cut-off at $|k| = 4.5$ for the storage data and $|k| = 3.5$ for the loss data. An initial 3 points differentiation is done for the storage data. The solid line, the circles and the plusses have the same meaning as in figure 7.4.

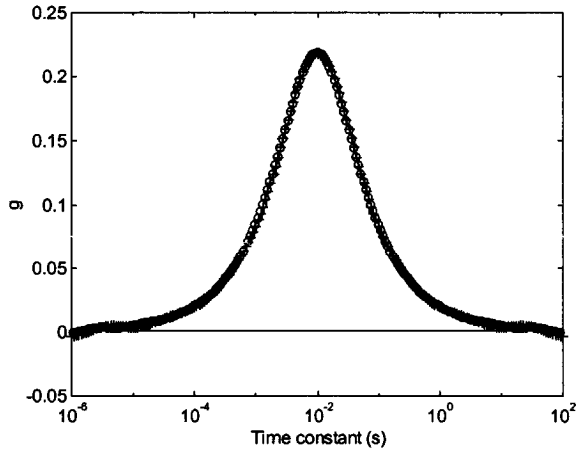


Figure 7.13. Deconvolution result for a Cole-Cole relaxation with $\nu = 0.6$. A zeroth order filter is used with cut-off at $|k| = 4.5$ for the storage data and $|k| = 3.0$ for the loss data. An initial 3 points differentiation is done for the storage data. The solid line, the circles and the plusses have the same meaning as in the previous figure.

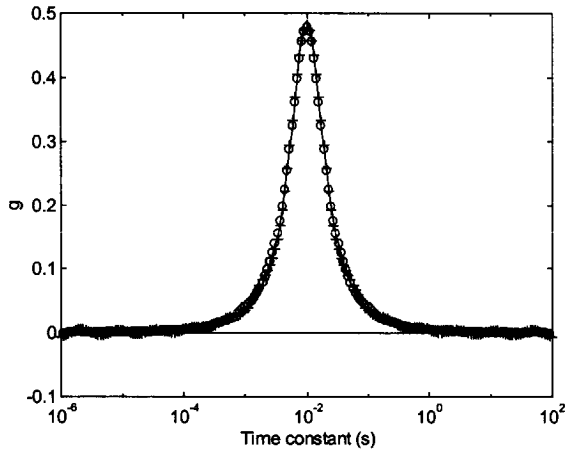


Figure 7.14. Deconvolution result for a Cole-Cole relaxation with $\nu = 0.8$. A zeroth order filter is used with cut-off at $|k| = 6.2$ for the storage data and $|k| = 5.0$ for the loss data. An initial 3 points differentiation is done for the storage data. The solid line, the circles and the plusses have the same meaning as in the previous figure.

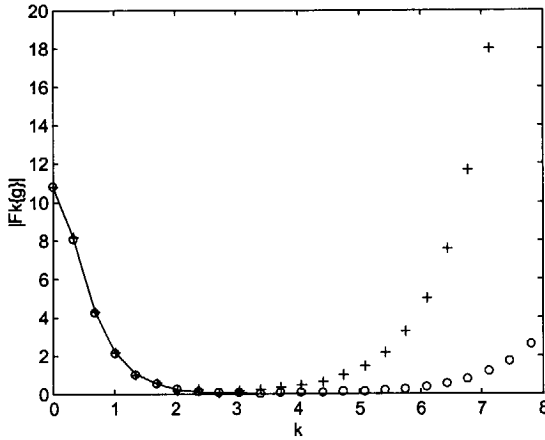


Figure 7.15. Unfiltered Fourier transform (modulus) of g for a Cole-Cole relaxation with $\nu = 0.6$. The solid line, the circles and the plusses have the same meaning as in the previous figure.

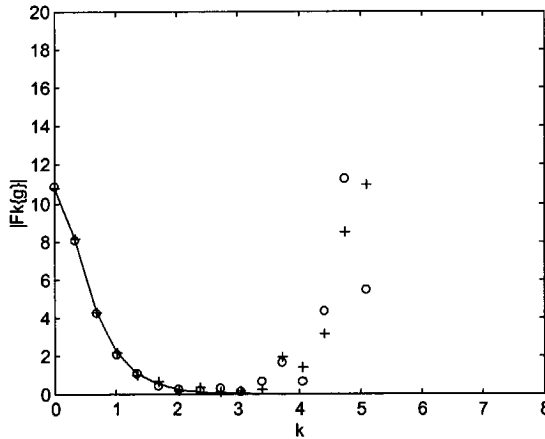


Figure 7.16. Unfiltered Fourier transform (modulus) of g for a Cole-Cole relaxation with $\nu = 0.6$. The solid line, the circles and the plusses have the same meaning as in the previous figure. The noise amplitude was 0.001.

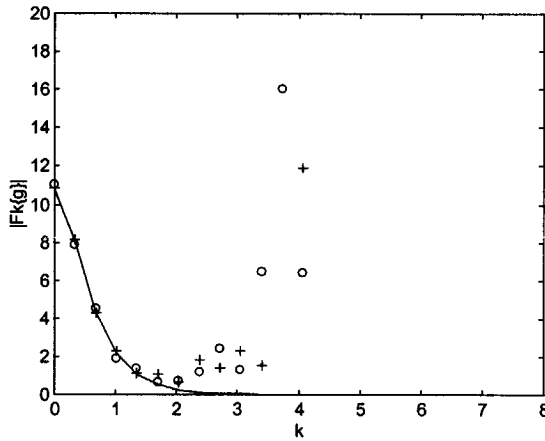


Figure 7.17. Unfiltered Fourier transform (modulus) of g for a Cole-Cole relaxation with $\nu = 0.6$. The solid line, the circles and the plusses have the same meaning as in the previous figure. The noise amplitude was 0.01.

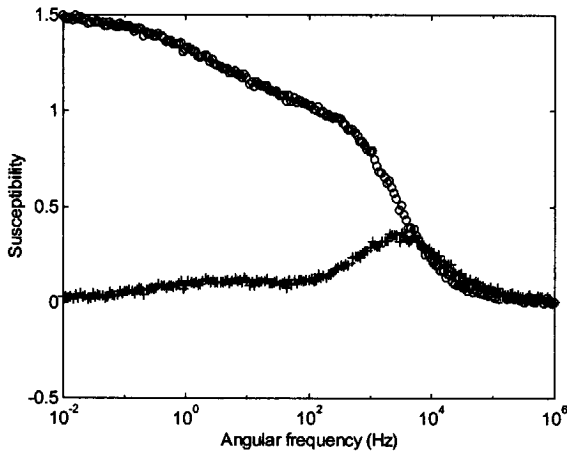


Figure 7.18. Test spectrum of two simultaneous Cole-Cole relaxations at angular frequencies 3Hz and 3000Hz, with Cole-Cole parameters 0.5 and 1.0 and amplitudes 0.5 and 1.0. The spheres show the real part and the plusses show minus the imaginary part of the susceptibility. The noise amplitude was 0.01.

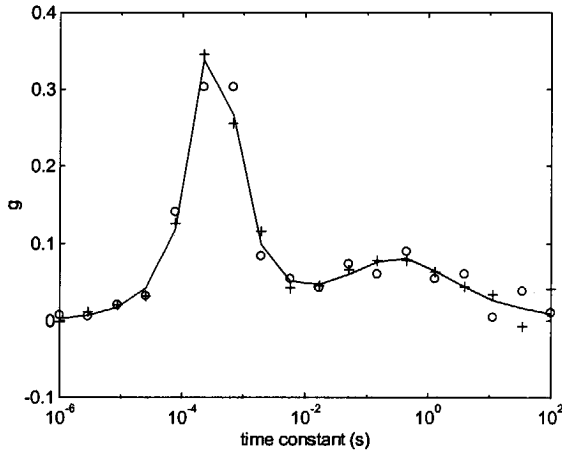


Figure 7.19. Least square fit of the test spectrum. The solid line shows the analytic result, the circles the inversion for the storage data, the plusses for the loss data.

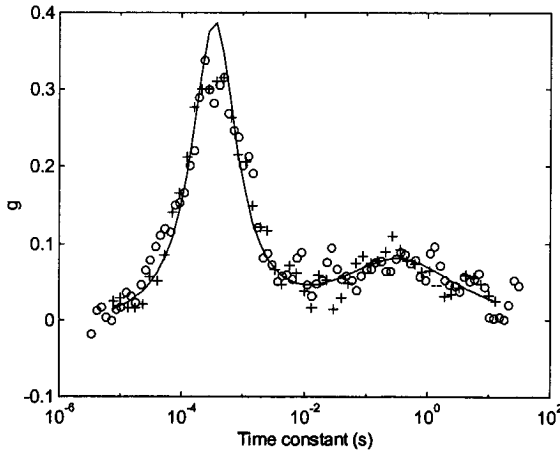


Figure 7.20. Differentiation result of the test spectrum, after 1 iteration for storage data and 2 iterations for loss data. The differentiation interval was 11 points wide and a 2 points decimation was done. The solid line, the circles and the plusses have the same meaning as in the previous figure.

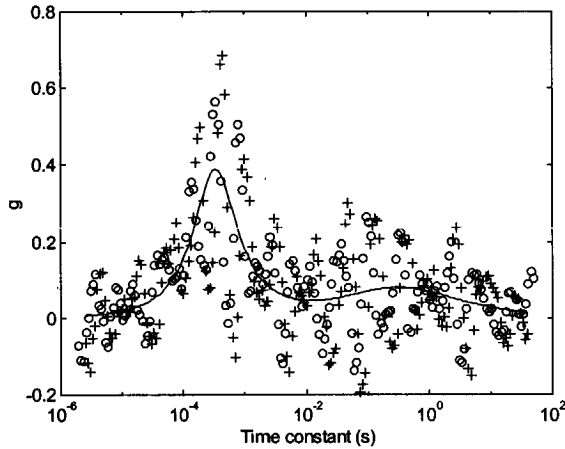


Figure 7.21. Differentiation result of the test spectrum after 1 iteration for storage data and 2 iterations for loss data. The differentiation interval was 11 points wide and no decimation was done. The solid line, the circles and the plusses have the same meaning as in the previous figure.

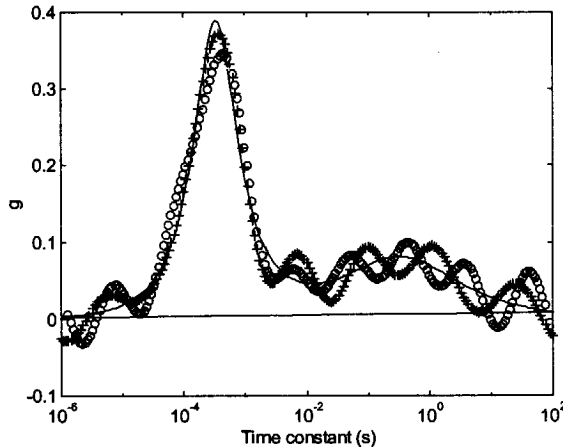


Figure 7.22. Deconvolution of test spectrum. A zeroth order filter was used with cut-off frequency $|k| = 3.0$, both for storage and loss data. An initial 7 points differentiation was done for storage data. The solid line, the circles and the plusses have the same meaning as in the previous figure.

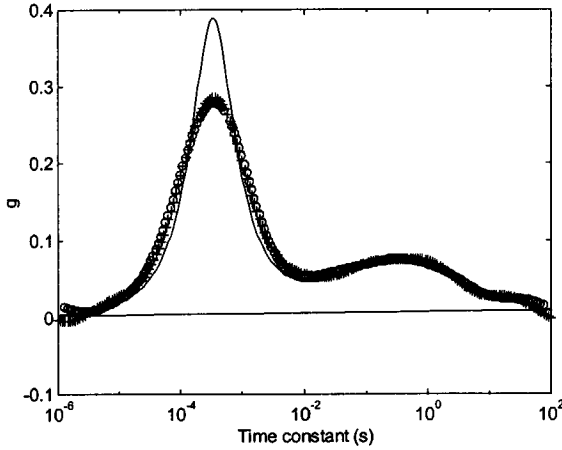


Figure 7.23. Deconvolution of test spectrum. A second order filter was used with cut-off frequency $|k| = 3.0$, both for storage and loss data. An initial 7 points differentiation was done for storage data. The solid line, the circles and the plusses have the same meaning as in the previous figure.

Appendix 7A. A factorized differential operator

Here we will show that the integral equations:

$$\chi'(\omega) = \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) \frac{1}{1 + (\omega\tau)^2} d \ln \tau \equiv \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) h'(\omega\tau) d \ln \tau \tag{7A.1a}$$

$$\chi''(\omega) = \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) \frac{\omega\tau}{1 + (\omega\tau)^2} d \ln \tau \equiv \int_{\ln \tau = -\infty}^{\infty} g(\ln \tau) h''(\omega\tau) d \ln \tau \tag{7A.1b}$$

are inverted by the following operators:

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} \prod_{n=\text{even}} \left(1 - \frac{D^2}{n^2}\right) [-D\chi'(\omega)] \tag{7A.2a}$$

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} \prod_{n=\text{odd}} \left(1 - \frac{D^2}{n^2}\right) \left[\frac{2}{\pi} \chi''(\omega)\right] \tag{7A.2b}$$

where D is shorthand for $d/d(\ln \omega)$. The proof is rather lengthy, but straightforward and follows the reasoning of Schwarzl and Staverman [Schwarzl 1953]. First let us consider the lowest order approximations according to (7A.2a) and (7A.2b):

$$g(\ln \frac{1}{\omega}) \cong -\frac{d\chi'}{d \ln \omega} = 2 \int_{\ln \tau = -\infty}^{\infty} [h''(\omega\tau)]^2 g(\ln \tau) d \ln \tau \quad (7A.3a)$$

$$g(\ln \frac{1}{\omega}) \cong \frac{2}{\pi} \chi''(\omega) = \frac{2}{\pi} \int_{\ln \tau = -\infty}^{\infty} h''(\omega\tau) g(\ln \tau) d \ln \tau \quad (7A.3b)$$

The idea behind this approximation is that the first and second order powers of $h''(\omega\tau)$ are regarded as approximations of the Dirac delta function $\delta[\ln(\omega\tau)]$ and replaced by it. This makes sense because the function:

$$h''(\omega\tau) = \frac{\omega\tau}{1 + (\omega\tau)^2} = \frac{1}{\cosh[\ln(\omega\tau)]}$$

is a symmetric, peaked function around $\tau = 1/\omega$ on the $\ln(\tau)$ scale. Therefore, any power of $h''(\omega\tau)$ can approximate the Dirac delta function $\delta[\ln(\omega\tau)]$, apart from a normalization factor. For the first and second powers, these normalization factors are:

$$Q_{11} = \int_{\ln \tau = -\infty}^{\infty} h''(\omega\tau) d \ln \tau = \frac{\pi}{2}$$

$$Q_{22} = \int_{\ln \tau = -\infty}^{\infty} h''(\omega\tau)^2 d \ln \tau = \frac{1}{2}$$

The reciprocals of Q_{11} and Q_{22} are the prefactors of the approximations (7A.3a) and (7A.3b). Now the way to higher order approximations is to construct integrals with sharper kernels than (7A.3a) and (7A.3b), which are better approximations of the Dirac delta function. This can be done by differentiations of the integral equations with respect to ω , which affects $\chi'(\omega)$ and $\chi''(\omega)$ as well as the kernels $h'(\omega\tau)$ and $h''(\omega\tau)$, but leaves $g(\ln\tau)$ intact. Since higher powers $[h''(\omega\tau)]^n$ are sharper kernels than $h''(\omega\tau)$ itself, it follows that what we need is an operator turning $h''(\omega\tau)$ into higher powers of itself. This can be done in different ways, but we will do it in such a way that a given power of $h''(\omega\tau)$ is turned into a higher power. More precisely, an operator will be constructed that turns $[h''(\omega\tau)]^n$ into $[h''(\omega\tau)]^{n+2}$.

We start by observing that by logarithmic differentiation of $h'(\omega\tau)$ and $h''(\omega\tau)$ up to any order, one always arrives at a linear combination of polynomial fractions of the following form:

$$\frac{[\omega\tau]^m}{[1 + \omega^2 \tau^2]^n} \equiv h_{mn}(\omega\tau) \equiv |m, n| \quad (7A.4)$$

This follows from the fact that $h'(\omega\tau) = h_{01}(\omega\tau) = |0, 1|$ and $h''(\omega\tau) = h_{11}(\omega\tau) = |1, 1|$ and from the following relation:

$$D|m, n\rangle = \omega \frac{d}{d\omega} \frac{[\omega\tau]^m}{[1 + \omega^2\tau^2]^n} = m|m, n\rangle - 2n|m + 2, n + 1\rangle$$

We see that by logarithmic differentiation, the higher order polynomial fraction $|m + 2, n + 1\rangle$ can be obtained. The result also contains the original polynomial fraction $|m, n\rangle$, but this contribution can be subtracted, which leads to the construction of a new operator A_{mn} :

$$A_{mn}|m, n\rangle \equiv \left(\frac{m - D}{2n}\right)|m, n\rangle = |m + 2, n + 1\rangle \tag{7A.5}$$

In addition, we note the following possibility to write a polynomial fraction $|m, n\rangle$ as the sum of two other polynomial fractions:

$$|m, n\rangle = \frac{x^m}{[1 + x^2]^n} = \frac{x^m[1 + x^2]}{[1 + x^2]^{n+1}} = |m, n + 1\rangle + |m + 2, n + 1\rangle$$

This is a very useful relation, since the second part also appears in the right hand side of (7A.5). It follows that the operator $B_{mn} = 1 - A_{mn}$ working on $|m, n\rangle$ also results in a single polynomial fraction:

$$B_{mn}|m, n\rangle \equiv (1 - A_{mn})|m, n\rangle = |m, n + 1\rangle$$

Hence the elements $|m, n\rangle$ form a closed group under the operations A_{mn} and B_{mn} . Now we observe the following:

$$A_{n, n+1}B_{nn}|n, n\rangle = |n + 2, n + 2\rangle \tag{7A.6}$$

This provides us the machinery to construct symmetric kernels $h_{nn}(\omega\tau) = |n, n\rangle$ of arbitrarily high order, starting at $h''(\omega\tau) = h_{11}(\omega\tau) = |1, 1\rangle$ for kernels with odd n and at $-Dh'(\omega\tau) = h_{22}(\omega\tau) = |2, 2\rangle$ for kernels with even n .

This is the key to higher order approximations of $g(\ln\tau)$ from either $\chi''(\omega)$ or $\chi'(\omega)$ data by only second order differentiations. The only problem is that the kernels $h_{nn}(\omega\tau)$ are not properly normalized for approximating the Dirac delta function. For that, they have to be divided by normalization factors Q_{nn} such that:

$$\frac{1}{Q_{nn}} \int_{\ln \tau = -\infty}^{\infty} h_{nn}(\omega\tau) d \ln \tau = 1$$

Closed expressions for the normalization factors can be given, even for the normalization factors Q_{mn} of the possibly asymmetric $h_{mn}(\omega\tau)$:

$$Q_{nm} \equiv \int_{\ln \tau = -\infty}^{\infty} h_{nm}(\omega\tau) d \ln \tau = \int_{x=0}^{\infty} h_{nm}(x) \frac{dx}{x} = \int_{x=0}^{\infty} \frac{x^{m-1}}{[1+x^2]^n} dx$$

This is done by comparison with the Beta-function:

$$B(\alpha, \beta) \equiv \int_{t=0}^1 t^{\alpha-1} (1-t)^{\beta} dt = 2 \int_{x=0}^{\infty} \frac{x^{2\beta-1} dx}{[1+x^2]^{\alpha+\beta}}$$

This leads to the following expression for the normalization factors:

$$Q_{nm} = \frac{1}{2} B\left(n - \frac{m}{2}, \frac{m}{2}\right) = \frac{\Gamma\left(n - \frac{m}{2}\right) \Gamma\left(\frac{m}{2}\right)}{2\Gamma(n)}$$

Now we can define the normalized kernels $\hat{h}_{nm}(\omega\tau)$:

$$\hat{h}_{nm}(\omega\tau) \equiv \frac{2\Gamma(n)}{\Gamma\left(n - \frac{m}{2}\right) \Gamma\left(\frac{m}{2}\right)} h_{nm}(\omega\tau)$$

of which the symmetric ones $\hat{h}_{nn}(\omega\tau)$ can approximate the Dirac delta function. The two lowest order examples appear in (7A.3a) and (7A.3b):

$$\hat{h}_{11}(\omega\tau) = \frac{2\Gamma(1)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right)} h_{11}(\omega\tau) = \frac{2}{\pi} \frac{\omega\tau}{[1+(\omega\tau)^2]}$$

$$\hat{h}_{22}(\omega\tau) = \frac{2\Gamma(2)}{\Gamma(1)\Gamma(1)} h_{22}(\omega\tau) = 2 \left(\frac{\omega\tau}{1+(\omega\tau)^2} \right)^2$$

The operators A_{mn} and B_{mn} can now be adjusted in such a way that the normalization of $\hat{h}_{mn}(\omega\tau)$ is conserved:

$$\hat{A}_{mn} \equiv \frac{Q_{nm}}{Q_{m+2, n+1}} A_{mn} = \frac{2n}{m} A_{mn} = \left(1 - \frac{1}{m} D\right)$$

$$\hat{B}_{mn} \equiv \frac{Q_{nm}}{Q_{m, n+1}} B_{mn} = \frac{2n}{2n-m} B_{mn} = \left(1 - \frac{1}{2n-m} D\right)$$

This leads to the following normalized version of the operator of (7A.6):

$$\hat{A}_{n, n+1} \hat{B}_{mn} = \left(1 - \frac{1}{n} D\right) \left(1 + \frac{1}{n} D\right) = \left(1 - \frac{D^2}{n^2}\right) \quad (7A.7)$$

This operator turns $\hat{h}_{nn}(\omega\tau)$ into $\hat{h}_{n+2, n+2}(\omega\tau)$:

$$\left(1 - \frac{D^2}{n^2}\right) \hat{h}_{nn}(\omega\tau) = \hat{h}_{nn}(\omega\tau) - \frac{1}{n^2} \frac{d^2}{d \ln(\omega)^2} \hat{h}_{nn}(\omega\tau) = \hat{h}_{n+2, n+2}(\omega\tau)$$

The operator (7A.7) allows us to stepwise improve the lowest order approximations (7A.3a) and (7A.3b). For that purpose, we first rewrite the integrals of (7A.3a) and (7A.3b) in terms of normalized kernels:

$$-D\chi'(\omega) = \int_{\ln \tau = -\infty}^{\infty} \hat{h}_{22}(\omega\tau) g(\ln \tau) d \ln \tau \tag{7A.8a}$$

$$\frac{2}{\pi} \chi''(\omega) = \int_{\ln \tau = -\infty}^{\infty} \hat{h}_{11}(\omega\tau) g(\ln \tau) d \ln \tau \tag{7A.8b}$$

Now by applying the operator $(1 - D^2/n^2)$ a number of times on these integrals, the kernels are sharpened to $\hat{h}_{nn}(\omega\tau)$ with higher n and hence, the left hand sides of the integrals gradually become better approximations of $g[\ln(1/\omega)]$. Here, n runs through the even numbers 2, 4, 6 and so on for (7A.8a) and through the odd numbers 1, 3, 5 and so on for (7A.8b). This may be written into the following compact form:

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} \prod_{n=\text{even}} \left(1 - \frac{D^2}{n^2}\right) [-D\chi'(\omega)] \tag{7A.9a}$$

$$g\left(\ln \frac{1}{\omega}\right) = \lim_{n \rightarrow \infty} \prod_{n=\text{odd}} \left(1 - \frac{D^2}{n^2}\right) \left[\frac{2}{\pi} \chi''(\omega)\right] \tag{7A.9b}$$

These are the final results. The factorized form of the inversion operators has the advantage that following approximations can be obtained from the previous by only a second order differentiation $(1 - D^2/n^2)$. The first step of the procedure leads to:

$$g_2\left(\ln \frac{1}{\omega}\right) = \left(1 - \frac{D^2}{2^2}\right) [-D\chi'(\omega)]$$

$$g_1\left(\ln \frac{1}{\omega}\right) = \left(1 - \frac{D^2}{1^2}\right) \left[\frac{2}{\pi} \chi''(\omega)\right]$$

By working out the differentiations, it can be easily verified that these expressions are identical to the highest order results given by Schwarzl and Staverman (cf. [Böttcher 1978]). For the inversion algorithm however, we do not work out the expressions and proceed a following second order differentiation:

$$g_4\left(\ln \frac{1}{\omega}\right) = \left(1 - \frac{D^2}{4^2}\right) g_2\left(\ln \frac{1}{\omega}\right)$$

$$g_3\left(\ln\frac{1}{\omega}\right) = \left(1 - \frac{D^2}{3^2}\right)g_1\left(\ln\frac{1}{\omega}\right)$$

and so on. In this way one can improve the result to arbitrarily high order approximations by only carrying out second derivatives.

References

- Aa, M. van der, *Wall Ironing of Polymer Coated Sheet Metal*, PhD Thesis, Eindhoven University of Technology (1999)
- Alfrey, T., Gurnee, E.F., Lloyd, W.G., *J. Polym. Sci. C* **12** (1966), 249
- Azzerri, N., *J. Electroanal. Chem.* **66** (1975), 99
- Bacon, R.C., Smith, J.J., Rugg, F.M., *Ind. And Eng. Chem.* **40** (1948), 161
- Baikie, I.D., Estrup, P.J., *Rev. Sci. Instrum.* **69** (1998), 3902
- Baumgärtner, H., *Meas. Sci. Technol.* **3** (1992), 237
- Beaunier, L., Epelboin, I., Lestrade, J.C., Takenouti, H., *Surf. Techn.* **4** (1976), 237
- Bellucci, F., Nicodemo, L., *Corrosion* **49** (1993), 235
- Bird, D.W., Symp. *Int. Corr. of Cans, Liège* **2** (1971), 149
- Blok, H., Oristaglio, M.L., *Wavefield Imaging and Inversion in Electromagnetics and Acoustics*, TU Delft, Report number: Et/EM 1995-21
- Bockris, J. O'M., Dražić, D.M., *Electro-Chemical Science*, Taylor & Francis Ltd., London (1972)
- Bockris, J. O'M., Reddy, A.K.N., *Modern Electrochemistry*, Plenum Press (1970)
- Boer, J.S.W. de, Krusemeyer, H.J., Burhoven Jaspers, N.C., *Rev. Sci. Instrum.* **44** (1973), 1003
- Bonora, P.L., Deflorian, F., Fedrizzi, L., Rossi, S., *Special Publication No 177*, The Royal Society of Chemistry (1988), 163
- Böttcher, C.J.F., Bordewijk, P., *Theory of Electric Polarization*, Elsevier Scientific Publishing Company (1978)
- Bovard, F.S., *Electrochimica Acta* **40** (1995), 201
- Boyce, W.E., DiPrima, R.C., *Elementary Differential Equations and Boundary Value Problems*, 3rd Ed. J. Wiley & Sons (1977)
- Brasher, D.M., Kingsbury, A.H., *J. Appl. Chem.* **4** (1954), 96
- Braune, H., *Z. Physikal. Chem.* **110** (1924), 147
- Bugnard, C., *Proc. Sixth Int. Tinplate Conf.*, London, (1996), 297
- Burt, R.B., Willis, J., *J. of Sci. Instrum.* **34** (1957), 177
- Butler, J.A.V., *Trans. Faraday Soc.* **19** (1924), 734
- Cahn, J.W., *Materials Research Society Symposium Proceedings Vol. 398*, Materials Research Society (1996), 425
- Chen, C., Breslin, C.B., Mansfeld, F., *Materials and Corrosion*, **49** (1998), 569
- Chen, Y.C., Cunningham, J.E., Flynn, C.P. *Phys. Rev.* **B30** (1984), 7317
- Cherry, B.W., *Australasian Corros. Eng.* **18** (1974), 23
- Cole, K.S., Cole, R.H., *J. Chem. Phys.* **9** (1941), 341
- Comyn, J., *Polymer Permeability*, Elsevier (1985)
- Courant, R., Hilbert, D., *Methods of mathematical physics*, Vol. 1 & 2, New York Interscience (1989)

- Craig, P.P., *Phys. Rev. Lett.* **22** (1969), 700
- Crank, J., *Diffusion in Polymers* (1968)
- Darowicky, K., *Corrosion Science* **37** (1995), 913
- Debye, P., *Phys. Zeit.* **13** (1912), 97
- Delucchi, M., Barbucci, A., Cerisola, G., *Electrochim. Acta* **44** (1999), 4297
- DeVecchio, D., Bhushan, B., *Rev. Sci. Instrum.* **69** (1998), 3618
- Dissado, L.A., Hill, R.M., *Nature* **279** (1979), 685
- Dolfing, J., *Mat. Sci. Forum* **192-194** (1995), 497
- Dubpennell, G., Westbrook, J.H., *Proceedings of the symposium on: 'Selected Topics in the History of Electrochemistry'* The Electrochemical Society, Proceedings Volume 78-6 (1978)
- Duderstadt, J.J., Hamilton, L.J., *Nuclear Reactor Analysis*, Wiley (1976)
- Ederle, E., Thesis U. der Bundeswehr München (1990)
- Einstein, A., *Ann. Physik* **4** (1905), 549
- Emri, I., Tschögl, N.W., *Rheologica Acta* **32** (1993), 311
- Erday-Grúz, T., Volmer, M., *Z. Physik. Chemie* **150A** (1930), 203
- Exalto, D.E., *The Scanning Kelvin Probe, New developments*, Final report 'TWAIO', Delft University of Technology, Laboratory of Materials Science (1998)
- Ferry, J.D. *Viscoelastic Properties of Polymers*, Wiley (1968)
- Feynman, R.P., *The Feynman Lectures on Physics*, Volume 1 (1963)
- Fick, A., *Pogg. Ann.* **94** (1855), 59
- Fourier, J.B.J., *Théorie analytique de la chaleur* (1822)
- Friedrich, C. *Rheologica Acta*, **30** (1991), 7
- Galic, K. *Electrochem. Soc. Proc.* **97-41** (1998), 161
- Garrison, B.J., Srinivasan, R., *Appl. Phys. Lett.* **44** (1984), 849
- Geenen, F.M., Westing, E.P.M. van, Wit, J.H.W. de, *11-th International Corrosion Congress*, Florence (1990), Paper 174
- Glass, A.L., Smith, J., *J. of Paint Technol.* **39** (1967), 490
- Glover, A.C.J., Illy, E.K., Piper, J.A., *IEEE J. of Selected topics in quantum electronics* **1** (1995), 830
- Goddard, P.J., Lambert, R.M., *Surf. Sci.* **67** (1977), 180
- Grandle, J.A., *Corrosion* **50** (1994), 793
- Hartman, L., *Case Studies in Manufacturing with Advanced Materials, Vol. 1*, Ed. by Wit, J.H.W. de, Demaid, A., Onillon, N. (1992), 247
- Hartshorn, L., Megson, N.J.L., Rushton, E., *J. Soc. Chem. Ind. London, Transactions* **56** (1937), 266T
- Herrmann, C.C., Perrault, G.G., Pilla, A.A., *Anal. Chem.* **40** (1968), 1173
- Hill, S., *Materials World*, February (1996), 69
- Hine, F., *J. of the Electrochemical Society* **103** (1956), 997
- Hirayama, R., Haruyama, S., *Corrosion* **47** (1991), 952
- Honerkamp, J., *J. Chem. Phys.* **98** (1993), 865
- Hosford, W.F., *Scientific American*, September (1994), 34
- Isaacs, H.S., *Corrosion* **36** (1980), 269
- Isaacs, H.S., Kendig, M.W., *Corr. Sci.* **28** (1988), 547
- Jaffe, L.F., Naticelli, R., *J. Cell Biology* **63** (1974), 614
- Jong, A.J. de, *Corrosion Behavior of Polymer Coated Packaging Steel*, First year report of a PhD study, TU Delft, Materials Science Laboratory (1997)

- Jong, A.J. de, *Corrosion Behavior of Polymer Coated Packaging Steel*, 2½ year report of a PhD study, TU Delft, Materials Science Laboratory (1998)
- Jonscher, A.K., *Dielectric Relaxation in Solids*, Chelsea Dielectrics Press (1983)
- Jonscher, A.K., *Nature* **253** (1975), 717
- Kamm, G.G., *ISU Internat.* **29** (1989), 614
- Kasai, T., Rigney, D.A., Zharin, A., *Scripta Materialia* **39** (1998), 561
- Kelvin, Lord W., *Phil. Mag.* **46** (1898), 82
- Kendig, M., Scully, J. *Corrosion* **46** (1990), 22
- Kikukawa, A., Hosaka, S., Imura, R., *Rev. Sci. Instrum.* **67** (1996), 1463
- Kittel, C. *Introduction to Solid State Physics* 5th Ed., J. Wiley & Sons (1976)
- Kittel, C., Kroemer, H., *Thermal Physics* 2nd Ed., W.H. Freeman and Company (1980)
- Kittelberger, W.W., Elm, A.C., *Ind. And Eng. Chem.* **44** (1952), 326
- Kleniewski, A., *Br. Corros. J.* **10** (1975), 91
- Kloppers, M.J., Bellucci, F., Latanision, R.M., Brennan, J.E., *J. of Appl. Polym. Sci.* **48** (1993), 2197
- Kohlrausch, R., *Pogg. Ann. Physik. Chemie* **91** (1854), 179
- Koot, E.J.H., *Masters Thesis*, TU Delft, Faculty of Materials Science (1991)
- Kubal, M., Panacek, F., *British Corr. J.* **30** (1995), 309
- Lasoski, S.W., Cobbs, W.H., *J. of Polym. Sci.* **36** (1959), 21
- Lauterbach, M.T., *Scanning Kelvin Probe-Metingen aan Beluchttingscellen*, Final report M.Sc. (1999)
- Lenderink, H.J.W., *Filiform Corrosion on coated Aluminum Alloys*, Thesis T.U. Delft (1995)
- Levie, R. de, *Electrochimica Acta* **8** (1963), 751
- Ließ, H.D., Mäckel, R., Ren, J., *Surface and Interface Analysis* **25** (1997), 855
- Light, R.R., Saymour, R.W., *Polymer Eng. And Sci.* **22** (1982), 857
- Lindqvist, S.A., *Corrosion-Nace* **41** (1985), 69
- Link, N., Bauer, S., Ploss, B. *J. Appl. Phys.* **69** (1991), 2759
- Lorrain, P., Corson, D.L., *Electromagnetic Fields and Waves*, W.H. Freeman and Company (1962)
- Love, E.R., Byrne, A., *J. London Math. Soc.* (1980), 285
- Macdonald, J.R., *Impedance Spectroscopy*, J. Wiley & Sons (1987)
- Mäckel, R., Baumgärtner, H., Ren, J., *Rev. Sci. Instrum.* **64** (1993), 694
- Maeres, P., *Trans. Faraday Soc.* **54** (1958), 40
- Makar, R., Boothroyd, A.R., Cherry, E.C., *Nature* **29** (1948), 845
- Mansfeld, F., *J. of Appl. Electrochemistry* **25** (1995), 187
- Mansfeld, F., Xiao, H., Han, L.T., Lee, C.C., *Progress in Org. Coatings* **30** (1997), 89
- Mayne, J.E.O., *Br. Corros. J.* **5** (1970), 106
- McCrum N.G., *Anelastic and Dielectric Effects in Polymeric Solids* (1967)
- McWhirter, J.G., Pike, E.R., *J. Phys. A*, **11** (1978), 1729
- Meijers, S.D., *The Kelvin Probe, Development of a scanning Kelvin probe for electrochemical applications*, Final report 'TWAIO', Delft University of Technology, Laboratory of Materials Science (1996)
- Meinke, W.W., *J. Chem. Phys.* **31** (1959), 978
- Mende, S., *PHYSIK, Gleichungen und Tabellen*, VEB Fachbuchverlag (1974)
- Menges, G., Schneider, W., *Kunststofftechnik* **12** (1973), 265, 316 and 343
- Michaels, A.S., Vieth, W.R., Barrie, J.A., *J. of Appl. Phys.* **34** (1963), 1
- Montenari, A., *Prog. Org. Coat.* **29** (1996), 614

- Murray, J.D., *J. Oil Col. Chem. Assoc.* **56** (1973), 210
- Myers, A.W., Rogers, C.E., Stannett, V., M. Szwarc, *Modern Plastics* **34** (1957), 157
- Nabhan, W., et al., *Rev. Sci. Instrum.* **68** (1997), 3108
- Nazarov, A. Thierry, D., *J. Electrochem. Soc.* **145** (1998), L39
- Nernst, W. *Z. Physikal. Chem.* **2** (1888), 613
- Nernst, W., *Zur Theorie umkehrbare Elementen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften (1889), 83
- Newman, P., *Tin International* **72** (1999), 14
- Oppenheim, A.V., Willisky, A.S., Young, A.T., *Signals and Systems*, Prentice/Hall International (1983)
- Parker, R.L., *Geophysical Inversion Theory*, Princeton U. Press (1994).
- Pezzani, A., *Mat. Sci. Forum* **289-292** (1998), 259
- Priestly, M.B., *Spectral Analysis and Time Series*, Academic Press (1981)
- Reddish, W., *Trans. Of the Faraday Soc.* **46** (1950), 459
- Ren, J., Thesis U. der Bundeswehr München (1995)
- Reyna, L.G., Sobehart, J.R., *J. Appl. Phys.* **76** (1994), 4367
- Richards, J.M., Geisow, J.C.H., *J. Sci. Instrum.* **41** (1964), 28
- Rietveld, G. *Surface Spectroscopic Studies of High-Temperature Superconductors*, Thesis, T.U. Delft (1993)
- Sakalauskas, S., Sodeika, A., *Rev. Sci. Instrum.* **69** (1998), 466
- Schaeffert, R.M., *Electrophotography* 2nd Ed., Focal Press (London, 1970)
- Schmickler, W., *Chem. Rev.* **96** (1996), 3177
- Schmutz, P., Frankel, G.S., *J. Electrochem. Soc.* **145** (1998), 2285
- Schwarzl, F., Staverman, A.J., *Appl. Sci. Res.* **A4** (1953), 127
- Sekine, I., Hayakawa, T., Negishi, T., Yuasa, M., *J. Electrochem. Soc.* **137** (1990), 3029
- Seminikhin, O.A., et al., *Electrochimica Acta* **42** (1997), 3321
- Seruga, M., *Z. Lebensm. Unters. Forsch.* **202** (1996), 308
- Slee, J.A., Orchard, G.A.J., Bower, D.I., Ward, I.M., *J. of Polym. Sci.* **27** (1989), 71
- Smythe, W.R., *Static and Dynamic Electricity*, 3rd Ed. (1968)
- Sodiek, M., *Transactions ISU* **28** (1988), 663
- Spoelstra, M.B., *Filiform Corrosion of Anodised Aluminum Alloys*, Thesis, T.U. Delft (1999)
- Srinivasan, R., Braren, B., Dreyfus, R.W., *J. Appl. Phys.* **61** (1987), 372
- Srinivasan, R., *Science* **234** (1986), 559-565
- Stern, M., Geary, A.L., *J. Electrochem. Soc.* **104** (1957), 56
- Stratmann, M., *Corrosion Science*, **27** (1987), 869
- Stratmann, M., Streckel, H., *Corrosion Science* **30** (1990), 681
- Stratmann, M., Streckel, H., Feser, R., *Corrosion Science*, **32** (1991), 467
- Tafel, J., *Z. Physik. Chemie* **34** (1900), 187
- Tahara, A., Kodama, T., *Corrosion Engineering*, **46** (1997), 877
- Thomas, N.L., Windle, A.H., *Polymer* **23** (1982), 529
- Tschoegl, N.W., *The Phenomenological Theory of Linear Viscoelastic Behavior – An Introduction*, Springer (1989)
- Uhlig, H.H., *Corrosion & Corrosion Control*, J.Wiley & Sons (1963)
- Vetter, K., *Electrochemische Kinetik* Springer-Verlag (1961)
- Vittel Rao, R., Yaseen, M. *Pigment and Resin Technol.* **7** (1978), 4
- Wagner, C., *J. of the Electrochemical Society* **98** (1951), 116
- Walker, P., *Paint Technol.* **31** (1967), 22

- Wang, J., Hou, B., *Chin. J. Oceanol. Limnol.* **15** (1997), 36
- Weast, R.C., *Handbook of Chemistry and Physics*, 56th edition (1975)
- Weese, J., *Comput. Phys. Commun.* **77** (1993), 429
- Westing, E.P.M. van, Ferrari, G.M., Wit, J.H.W. de, *Corros. Sci.* **36** (1994), 979
- Widder, D.V., *The Laplace Transform*, Princeton U. Press (1946)
- Willems, C.R.J., *A Dielectric Study of Melting and Crystallization of Semi-rigid and Flexible-chain Polymers*, Thesis, T.U. Delft (1995)
- Williams, G., Watts, D.C., *Trans. Of the Farad. Soc.* **67** (1971), 1323
- Wind, M.M., Lenderink, H.J.W., *Progress in Org. Coat.* **28** (1996), 239
- Wit, J.H.W. de, *Corrosion Mechanisms in Theory and Practice*, Chapter 16, Ed. by Marcus, P. and Oudar, J. (1995)
- Wormwell, F., Brasher, D.M., *J. Iron St. Inst.* **162** (1949), 129
- Yasuda, H., Stannett, V., *J. of Polym. Sci.* **57** (1962), 907
- Zangwill, A. *Physics at Surfaces*, Cambridge U. Press (1988)
- Zisman, W.A., *Rev. Sci. Instrum.* **3** (1932), 367
- Znotins, T.A., Poulin, D., Reid, J., *Laser Focus/Electro-Optics* (may 1987), 54



Summary

Characterization of coatings inside food and beverage cans

At present, commercial food and beverage cans are produced from metal sheet first and coated afterwards. A new development, initiated at Corus RD&T, is to apply the coating first on the metal sheets and then to form the can from the coated sheets. This approach, called the “prefinished approach”, means a considerable simplification of the production process.

In order to further optimize the parameters of the prefinished approach, the performance of the coating inside the can must be tested. The task of the coating is to protect the metal of the can from corrosion and thus to prevent metal uptake by the stored product during a certain time, which is the shelf life of the stored product. Testing the corrosion protection of the coating can be done in a straightforward manner, which is to store a product in a can and to measure the uptake during shelf life. However, this procedure is time consuming and for the optimization of the production process, more rapid testing methods are desirable. In this thesis, a number of such methods is studied and applied to PET coated ECCS, which is one of the main composite materials for the prefinished approach.

The PET coatings provide corrosion protection of metals in two ways. The first is to block the transport of water and ions to the metal, and the second is to prevent the growth of microscopic defects by means of a good adhesion to the metal. Hence, the corrosion protection of a coating is characterized by the speed of water and ion transport through the coating and by the adhesion of the coating to the metal. In chapter 2, the transport of water and ions through PET is studied and in chapter 3, the adhesion of PET on ECCS.

The water transport through PET was found to obey Fick’s laws of diffusion. The transport of ions was so slow that only an upper limit for the diffusion coefficient could be given. The transport measurements were basically done by measuring the changes of the capacity of the PET coating during water and ion uptake. In addition some gravitational uptake measurements were done, where the mass increase of the PET coating was followed during water uptake. This provided a second, independent method for measuring the transport properties of PET, which was interesting for comparison and calibration. The two types of measurements agreed on the diffusion constant but a difference of a factor 2 was observed in the total water uptake of the coating.

Buffered citric acid of pH = 3 was the basic modeling electrolyte for consumables, but also sulfuric acid of pH = 1, 2, 3, 4 and phosphoric acid of pH = 2.5 also used. This however all lead to the same water transport. Also the influence of a 5% ethanol content was studied, modeling beer. It was expected that the wetting of PET would be improved by the ethanol content and thus that the transport properties of PET would be enhanced, but this has not been observed.

Also the influences of several heat treatments on the transport properties of PET were studied. This was done because several products receive an initial heat treatment just after they are stored in the can. The heat treatments were therefore done while the coating was immersed in citric acid. Heat treatments were studied at 75°C (pasteurization) and at 121°C and 2 bar (sterilization), for 15 and 60 minutes in both cases. However, no influence on the transport properties of PET was measured.

Finally, some long-term measurements indicated that PET coatings are not noticeably affected by phosphoric acid of pH = 2 for at least half a year. The slow water uptake, which is typical for hydrophilic polymers and which often leads to their breakdown, therefore does not seem to occur for the hydrophobic PET coatings.

For studying the adhesion of PET on ECCS under the action of an electrolyte, we worked with PET coated ECCS samples with artificial defects. Then by following the enlargement of the defect with time, we could follow the delamination rate of the coating. The size of the delaminated area was measured by electric impedance spectroscopy.

Crucial for this method is to work with well-defined defects having a well-defined shape and edges, such that the measurements can be reproduced and mutually compared. A rapid and easy way to create artificial defects is to puncture a needle with a fine tip through the coating until the needle makes electric contact with the metal. However, in this way the coating is significantly deformed at the edges of the defect and the adhesion of the coating on the metal may be affected. Another way of producing artificial defects is to use a laser. Probably the best suited are UV-lasers with a wavelength below about 300nm, since below this wavelength about all of the laser light is absorbed in the coating, and the heating of the metal underneath is as low as possible. Several types of Excimer-lasers qualify for the job, but also a YAG-laser, in case the fourth harmonic is used. However, due to the poor beam quality of such lasers, specialized optical equipment is required to produce well-defined defects. Future research is necessary to decide about laser defects.

Punctured artificial defects in PET were also used to study the deformation of the coating during the formation of the can. These measurements are discussed in chapter 4. It was observed that initially round defects become elliptical during the can formation process. The longer axis of the ellipses was always parallel with the axis of the can. In addition, the deformation of the defects was larger for defects higher in the can.

The observed deformations were in agreement with a simple analytical model, although the degree of deformation was significantly larger than predicted. In fact, the defects were closed after the complete can formation process. This was a surprising result, implying that initially present defects disappeared due to the prefinished approach.

We also considered several more advanced methods with the ambition of more locally studying corrosion phenomena. These methods, the SRET (Scanning Reference Electrode Technique), the SVET (Scanning Vibrating Electrode Technique) and the SKP (Scanning Kelvin Probe) are all electric potential scanners and therefore it was studied what can be derived from a scan of the electric potential in terms of local corrosion rates. The results of this study are presented in chapter 5.

The most versatile potential scanner is the SKP, by which also local adhesion differences of coatings can perhaps be measured. Therefore the attention was focused on the SKP, and chapter 6 is devoted to a study of this apparatus. The theoretical description of the

SKP measurements was studied as well as several ways to control the distance between the probe and the surface of the studied substrate. We found that the accuracy of the distance control system of the commercial SKP of UBM could be improved considerably by working with several voltages instead of with only one. The main reason for the improvement was that by our method, the distance control system is no longer dependent on potential variations on the studied surface.

Further improvement of the apparatus could be achieved by a laser distance control system. For that purpose, we proposed a combined probe and waveguide, which could be an optical fiber with a metal layer sputtered on it. By etching a lens on top of the fiber, an autofocus distance control system could be implemented. Then for instance the autofocus system of CD-players could be used to control the distance.

The measurements of chapter 2 did not give much information on the changes of the coating during heat treatments or aging. This was partly because the impedance measurements were carried out at only one frequency. Possibly from a permittivity spectrum of the coating, more information could be obtained. An elegant interpretation of permittivity spectra is to describe them by a time constant distribution. These time constants are the time constants of the dipole relaxations inside the polymer and describe the mobility of the main polymer chain and its side groups. The key is that these time constants are sensitive to changes of the polymer structure, and thus can be used to monitor these changes.

However, deduction of the time constant distribution from a permittivity spectrum requires the inversion of a Fredholm integral equation of the first kind. In chapter 7, three different methods were studied for doing this. These methods are the least square method, inversion by differential operators and deconvolution by Fourier transforms. All three methods were implemented in operational inversion algorithms by using Matlab[®], and all led to reasonably good results provided that the permittivity data were sufficiently "complete".

For the differential operator method, a new differential operator was constructed. This operator could be implemented into a general algorithm with higher order approximations with relative ease. Another apparently new result is a compact formula by which the Kramers-Kronig transforms could be carried out via Fourier transforms. Using this formula avoids the calculation of an integral with a pole on the integration interval, which is the case for the usual way of evaluating the Kramers-Kronig transforms.



Samenvatting

Karakterisering van coatings in conservenblikken

De productie van de conservenblikken die vandaag de dag in de handel zijn, vindt plaats door de blikken eerst van metalen platen te vormen en daarna met een beschermende coating te bekleden. Een nieuwe, bij Corus RD&T gefinitieerde ontwikkeling in het productieproces is om de coating eerst op de vlakke metalen platen aan te brengen alvorens hieruit de blikken te vormen. Deze ontwikkeling wordt de "prefinished approach" genoemd en betekent een aanzienlijke vereenvoudiging van het productieproces van conservenblikken.

Om het proces verder te optimaliseren moet de kwaliteit van de coating getest worden. De functie van de coating is om het metaal van het blikje tegen corrosie te beschermen en aldus de opname van metaalionen in het verpakte product te voorkomen. Het testen van de corrosiebescherming van een coating kan op een recht toe recht aan manier gedaan worden door een bepaald product in het blikje te verpakken en te controleren of er na verloop van tijd metaalionen in het product terecht komen. Is dat niet het geval gedurende de houdbaarheidsperiode van het product, dan geeft de coating voldoende bescherming. Deze methode kost echter veel tijd en in de ontwikkelingsfase van een productieproces is het handig om over snellere testmethoden te beschikken. In dit proefschrift zijn een aantal van zulke methoden onderzocht en toegepast op ECCS met een PET coating, wat een van de composietmaterialen is die onderzocht worden voor de "prefinished approach".

De PET coatings beschermen metalen op twee manieren tegen corrosie. Ten eerste vormt de coating een barrière voor water en ionen en ten tweede voorkomt de coating de groei van microscopische defecten door een goede hechting met het metaal. De corrosiebeschermende werking van een coating kan dus gekarakteriseerd worden door de snelheid waarmee water en ionen door de coating getransporteerd worden en de mate van hechting met het onderliggende metaal. In hoofdstuk 2 is het transport van water en ionen door PET onderzocht en in hoofdstuk 3 de hechting van PET op ECCS.

Het watertransport door PET bleek goed beschreven te kunnen worden met de diffusiewetten van Fick. Het ionentransport bleek zo langzaam dat niet meer dan een bovengrens voor de diffusiecoëfficiënt kon worden gegeven. De transportmetingen zijn voornamelijk gedaan middels de verandering van de capaciteit van de coating gedurende de opname van water en ionen. Daarbij zijn vergelijkende metingen gedaan door tevens het gewicht van de coating te volgen. Deze twee typen metingen kwamen goed overeen voor de diffusiecoëfficiënt, maar niet wat betreft de totale wateropname, waar een verschil van bijna een factor 2 werd waargenomen.

Voor de metingen is gebufferd citroenzuur van $\text{pH} = 3$ gebruikt als modelelektrolyt, maar ook zijn enige metingen gedaan in zwavelzuur van $\text{pH} = 1, 2, 3, 4$ en in fosforzuur van

pH = 2.5. Hierbij werd echter steeds dezelfde diffusieconstante van water gemeten. Ook is de invloed van 5% ethanol in de elektrolyt gemeten, zoals in bier. De verwachting was dat het ethanol de bevochtiging van PET zou verbeteren en daarmee de transporteigenschappen, maar dit is niet waargenomen.

Verder is de invloed van diverse warmtebehandelingen op het transportgedrag van PET onderzocht. Dit is gedaan omdat een aantal levensmiddelen een warmtebehandeling ondergaan nadat ze in de verpakking zijn gesloten. De warmtebehandelingen zijn daarom gedaan met een laag citroenzuur op de coating. Bestudeerd zijn warmtebehandelingen bij 75°C (pasteurisatie) en bij 121°C en 2 bar (sterilisatie), voor 15 en 60 minuten in beide gevallen. Opnieuw werd echter geen invloed op het transportgedrag van PET gemeten.

Tenslotte gaven langduriger metingen aan dat PET coatings minstens een half jaar lang niet aangetast wordt door citroenzuur. De langzame wateropname die kenmerkend is voor veel hydrofiele polymeren en die uiteindelijk ook voor falen van veel hydrofiele coatings zorgt, treedt dus niet op bij het hydrofobe PET.

Om de hechting van PET aan ECCS te meten is gewerkt met samples waarbij kunstmatige defecten in de PET coating waren aangebracht. Door de vergroting van de defecten te volgen als functie van de tijd onder citroenzuur, kon aldus de snelheid van delaminatie gevolgd worden, waarmee een maat voor de hechting van PET op ECCS is citroenzuur kon worden verkregen. De grootte van de defecten is gevolgd met impedantie spectroscopie.

Cruciaal voor deze methode is dat de kunstmatige defecten een goedgedefinieerde vorm en randen hebben, zodat de metingen reproduceerbaar zijn en onderling vergelijkbaar. Een snelle en eenvoudige manier om kunstmatige defecten te maken is door gaten in de coating te prikken met een naald totdat deze elektrisch contact maakt met het onderliggende metaal. Deze manier levert echter defecten op met vervormde randen, zodat de gemeten vergroting van de defecten wellicht geen goede maat is voor de werkelijke hechting van PET op ECCS. Een andere manier om kunstmatige defecten te maken is door een laser te gebruiken. Het best geschikt hiervoor zijn UV lasers met een golflengte lager dan 300nm, aangezien licht beneden deze golflengte zeer sterk door polymeren geabsorbeerd wordt. Hierdoor is de verhitting van het metaal bij het maken van een defect minimaal en daarmee de kans op onbeschadigde randen het grootst. Een aantal Excimer lasers komt in aanmerking, maar ook een YAG laser, indien de vierde harmonische gebruikt wordt. Vanwege de matige bundelkwaliteit van al deze lasers zijn wel speciale optische componenten vereist. Toekomstig onderzoek zal uitwijzen of laserdefecten inderdaad beter geschikt zijn.

Geprikte defecten in PET zijn ook gebruikt om de vervorming van PET te meten gedurende het productieproces van een blikje. Deze metingen zijn te vinden in hoofdstuk 4. Waargenomen is dat aanvankelijk ronde defecten ellipsvormig werden gedurende het vormingsproces. De lange as van de ellipsen was hierbij altijd evenwijdig met de as van de blikjes. Daarbij was de vervorming van de defecten groter naarmate de defecten verder van de bodem verwijderd waren. De gemeten vervormingen waren in overeenstemming met een eenvoudig analytisch model, hoewel de mate van vervorming vaak wel duidelijk groter was. Tijdens de laatste fase van het vormingsproces werden de defecten zelfs gesloten, zodat de "prefinished approach" in feite een helende werking op de PET coating had.

Verder zijn een aantal meer geavanceerde methoden bestudeerd, met de ambitie om corrosieprocessen meer lokaal waar te kunnen nemen. Deze methoden, te weten de SRET

(Scanning Reference Electrode Technique), de SVET (Scanning Vibrating Electrode Technique) en de SKP (Scanning Kelvin Probe) zijn alle elektrische potentiaal scanners en daarom is enige aandacht besteed aan de vraag wat men uit een potentiaalscan kan afleiden in termen van lokale corrosieactiviteiten aan een oppervlak. De bevindingen van deze studie zijn opgeschreven in hoofdstuk 5. De voornaamste conclusie is dat er geen duidelijke relatie bestaat tussen de corrosiepotentiaal en het corrosietempo. Daarbij wordt de elektrische potentiaal boven een oppervlak door diverse andere processen mede bepaald. Het meest geschikt lijkt de SKP, waarmee wellicht ook lokale verschillen in de hechting van de coating gemeten kunnen worden.

In hoofdstuk 6 wordt de SKP nader bestudeerd. In dit hoofdstuk wordt een model van de SKP metingen gegeven en worden diverse manieren besproken om de hoogte van de probe boven het substraat te regelen. Gevonden werd dat het hoogteregelingssysteem van de commerciële SKP van UBM een stuk nauwkeuriger gemaakt kon worden door niet met één, maar met een reeks spanningen de hoogte te regelen. De verbetering komt vooral door het feit dat de hoogteregeling zo niet meer van de potentiaalvariaties op het oppervlak afhangt.

Verdere verbetering van het apparaat zou bereikt kunnen worden met een laser hoogteregeling. Hierbij is bedacht om een optische fiber te gebruiken waarop een metallische laag gesputterd is. Aldus ontstaat een probe waardoor een lichtbundel naar het oppervlak geleid kan worden. Door een lens aan het eind van de fiber te etsen zou een autofocus systeem, zoals dat van een CD-speler, gebruikt kunnen worden om de hoogte te regelen.

De metingen van hoofdstuk 2 gaven niet veel inzicht in de wellicht kleine veranderingen in PET gedurende processen als warmtebehandelingen en veroudering. Meer informatie is wellicht te halen uit metingen van de permittiviteit als functie van de frequentie. Een elegante interpretatie van permittiviteitsspectra is om deze te beschrijven met tijdconstante verdelingen. Deze tijdconstanten zijn de tijdconstanten van de dipoolrelaxaties in het polymeer en beschrijven de beweeglijkheid van de hoofdpolymeerketen en zijn zijketens. De tijdconstanten zijn derhalve gevoelig voor veranderingen van de polymeerstructuur.

Het bepalen van de tijdconstanten uit dielectrische spectra is echter niet triviaal en vereist de inversie van een Fredholm integraalvergelijking van de eerste soort. In hoofdstuk 7 worden drie manieren met elkaar vergeleken om dit te doen. Dit zijn de kleinste kwadraten methode, inversie door differentiaaloperatoren en deconvolutie via Fourier transformaties. Al deze methoden zijn geïmplementeerd in werkende computerprogramma's met behulp van Matlab[®] en al deze methoden leidden tot redelijk goede resultaten, mits de data voldoende "compleet" waren.

Voor de differentiaaloperator methode is een nieuwe differentiaaloperator bedacht, welke vrij gemakkelijk geïmplementeerd kon worden in een algoritme voor hogere orde benaderingen. Een ander blijkbaar nieuw resultaat is een compacte formule waarmee de Kramers-Kronig transformaties uitgevoerd kunnen worden via Fourier transformaties. Hiermee wordt voorkomen dat er een integraal met een pool op het interval moet worden berekend, zoals bij de klassieke uitvoering van de Kramers-Kronig transformaties.



List of Publications

A.J. de Jong, *Corrosion behavior of polymer coated packaging steel, 1st year report of a PhD-study*. Rapport betreffende TU Delft projectnr. STM 94-209 aan: Koninklijke Hoogovens, IJmuiden, February 1997, 32 pp.

A.J. de Jong, *Corrosion behavior of polymer coated packaging steel, 1½ year report of a PhD-study*. Rapport betreffende TU Delft projectnr. STM 94-209 aan: Koninklijke Hoogovens, IJmuiden, June 1997, 12 pp.

A.J. de Jong, D.H. van der Weijde and J.H.W. de Wit, *EIS on well-defined defects in polymer coated steel*. in: O.R. Mattos (ed.), Proceedings of the 4th International Symposium on Electrochemical Impedance Spectroscopy (EIS 98), Angra dos Reis (Rio de Janeiro), Brazil, August 2nd – 7th 1998, pp. 204-206. COPPE/UFRJ, Rio de Janeiro, Brazil

A.J. de Jong and D.H. van der Weijde, *Kelvin probe measurements on coated metals*, in: Proceedings of EUROCORR '98: "Solutions to Corrosion Problems", Utrecht, September 28th – October 1st 1998, Section 6, 6 pp. Utrecht, Jaarbeurs (1998)

J.H.W. de Wit, D.H. van der Weijde, A.J. de Jong, F. Blekkenhorst and S.D. Meijers, *Local measurements in electrochemistry and corrosion technology*, in: P.L. Bonora and F. DeFlorian (eds.), Proceedings of the 6th International Symposium on Electrochemical Methods in Corrosion Research (EMCR VI), Trento, Italy, 25-29 August 1997. Materials Science Forum Vol. 289-292. Zürich, Switzerland, TransTech Publications Ltd. (1998) pp. 69-75. ISBN 0-87849-820-6

A.J. de Jong, *A one-dimensional model of Kelvin Probe measurements on coated metals*, Rapport betreffende TU Delft projectnr. STM 94-209 aan: Koninklijke Hoogovens, IJmuiden, April 1998, 26 pp.

A.J. de Jong, *Report on the 4th International Symposium on EIS, Rio de Janeiro, Brazil, August 2-7, 1998*. Rapport aan Koninklijke Hoogovens, IJmuiden, August 1998, 11 pp.

A.J. de Jong, *Making artificial defects in polymer coated steel*, Rapport betreffende TU Delft projectnr. STM 94-209 aan Koninklijke Hoogovens, IJmuiden, September 1998, 13 pp.

A.J. de Jong, *Corrosion behavior of polymer coated packaging steel, 2½ year report of a PhD-study*. Rapport betreffende TU Delft projectnr. STM 94-209 aan: Koninklijke Hoogovens, IJmuiden, October 1998, 14 pp.

A.J. de Jong, *Deformatie van PET op ECCS bij dieptrekken en wandstrekken*, Rapport betreffende TU Delft projectnr. STM 94-209 aan: Koninklijke Hoogovens, IJmuiden, November 1998, 12 pp.

A.J. de Jong, *Application of linear response theory to scanning probes*, Rapport betreffende TU Delft projectnr. STM 94.209 aan: Koninklijke Hoogovens, IJmuiden, December 1998, 15 pp.

A.J. de Jong, D.H. van der Weijde, D. Exalto and J.H.W. de Wit, *Kelvin probe measurements on coated metals*, Book of Abstracts EUROCORR '98, Utrecht, September 28th – Oktober 1st, p. 141.

A.J. de Jong, *Electrochemical impedance spectroscopy on artificial defects in PET coated ECCS*, Rapport betreffende TU Delft projectnr. STM 94.209 aan: Koninklijke Hoogovens, IJmuiden, February 1999, 15 pp., BTA/Rap, OZS/Inst MIDEG

A.J. de Jong, *Acoustic microscopy for corrosion studies on PET coated ECCS*, Rapport betreffende TU Delft projectnr. STM 94.209 aan: Koninklijke Hoogovens, IJmuiden, March 1999, 12 pp., BTA/Rap, OZS/Inst MIDEG

Oral presentations

D. Exalto, A.J. de Jong en J.H.W. de Wit, Development and testing of a Kelvin probe for the study of localized corrosion, Tweedaagse bijeenkomst van de SON Discussiegroep Elektrochemie, Lunteren, October 27th–28th 1997

A.J. de Jong, D.H. van der Weijde and J.H.W. de Wit, *EIS on well defined defects in polymer coated steel*, 4th International Symposium on Electrochemical Impedance Spectroscopy (EIS 98), Angra dos Reis (Rio de Janeiro), Brazil, August 2nd – 7th 1998

A.J. de Jong and D.H. van der Weijde, *Kelvin probe measurements on coated metals*, EUROCORR '98, Utrecht, September 28th – October 1st 1998

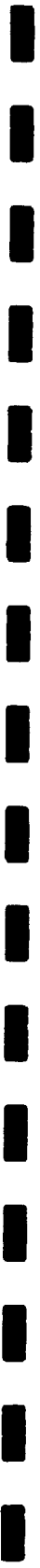
Poster presentations

A.J. de Jong, *Corrosion behavior of polymer coated packaging steel*, AIO Workshop "Non linear optics", Ameland, 9-13 juni 9th – 13th, 1997

A.J. de Jong, Distance control systems for scanning probes, Tweedaagse bijeenkomst Werkgemeenschap Chemie van de Vaste Stof en Materiaalkunde, Lunteren, januari 26th – 27th 1998.

Curriculum Vitae

The author of this thesis was born in Delft on may 6th 1965. He attended the high school *Erasmus College* in Zoetermeer from 1977 to 1983. In 1983 he started a MSc-study of applied physics, which he finished in may 1989. The topic of his MSc thesis was to design and build an experimental setup for measuring the polarization of γ -radiation, as well using this setup for measurements on graphite with implanted ^{160}Tb and graphite with intercalated $^{182}\text{Ta}_2\text{Cl}_5$. From 1990 to 1994 he worked as Onderzoek Project Assistant at the Interfaculty Reactor Institute of the TU Delft. Here he worked on physical models and computer simulations for evaluation of the power distribution in the core of nuclear power reactors. With these methods, a study on the optimal position of fuel elements in nuclear power reactors was carried out. After working for several temporary employment agencies in 1995, he started his PhD study in 1996. This study is described in this thesis.



Dankwoord

Tijdens mijn promotie heb ik het genoegen gehad met vele mensen samen te werken en een aantal daarvan wil ik hier op deze plaats bedanken:

- Hans de Wit, mijn promotor, die ondanks zijn drukke werkzaamheden bij Corus mijn onderzoek regelmatig met scherpzinnige commentaar bijgestuurd heeft;
- Jan van Turnhout, die altijd klaar stond voor “brainstormsessies”, die vaak pas spontaan laat in de middag begonnen en niettemin nog uren doorliepen. Hier kwamen vele ideeën uit voort en collega Michael Wübbenhorst voegde daar nog weer eens een groot aantal aan toe. Helaas kon slechts een beperkt deel hiervan worden uitgewerkt;
- Joseph Fehribach, wiens “sabbatical year” samenviel met het laatste jaar van mijn promotie, wat voor mij een groot geluk was. Hij was een vraagbaak voor theoretische problemen en leverde snel en deskundig commentaar op mijn teksten. Daarbij was hij een prettige kamergenoot die over elk onderwerp kan discussiëren;
- Hans Blok, van wie ik veel geleerd heb over inversietechnieken. Sessies bij hem waren altijd leerzaam, waarbij ik tot op de bodem werd uitgevraagd over mijn onderzoek;
- Hans Dieter Liess, die mij veel geleerd heeft over geavanceerde technieken betreffende de Kelvin Probe, veelal via email-discussies en via een zeer interessant gastcolloquium;
- Gerrit Bakker, bij wie ik altijd kon aanlopen met vragen over elektrochemie;
- Kas Hemmes en Richard Koops die veel meegedacht hebben over met name het experimentele werk;
- Mijn contactpersonen bij Corus, te weten Jacques Wijenberg, Jan Dolfing, Wim van Veenen, Conny Verhagen en Hans van der Weijde, die mij vaak geholpen hebben met goede en kritische adviezen. Daarnaast Paul Bongaerts, Henk Ras en René van Rijn voor het maken van gedeformeerde samples om aan te meten;
- Erik van Westing, die mijn teksten al vanaf de eerste versies van commentaar heeft voorzien en altijd bereid was te zeggen dat het allemaal wel goed zou komen. Ook Huib van Rooijen heeft dit enige tijd gedaan;
- De TWAIO’s Marjolein Deutekom en Steven Meijers, van wie ik vooral in mijn beginperiode veel geleerd heb. Later bij Corus heeft Steven mij ook nog vaak geholpen;
- Gert Rietveld, voor de vele adviezen bij het bouwen aan en meten met de Kelvin Probe;
- Marcel Loose, voor de adviezen op het gebied van signaalverwerking;
- Arie de Jong van FEL-TNO en Gon Weijers voor de infraroodmetingen aan coatings;
- Danny Exalto, met wie ik diens gehele TWAIO-periode heb mogen samenwerken. Lang niet alles wat we opgestart hebben is gerealiseerd, maar we hebben ervan geleerd en we hebben plezier gehad, zoals bijvoorbeeld toen we onder de naam “Prof Hans” met een gewaagd sprintersteam deelnamen aan de Tourtoto. Hans, in alles geïnteresseerd, doch niet in sport, was niettemin zo sportief ons de eerste prijs te laten;

- Wouter Hamer, die mij veel geholpen heeft met zijn kennis van materiaalkunde en computers. Dit geldt ook voor Siu Fai Au en Jeroen van den Brand. Het doet me plezier dat Jeroen een deel van mijn werk voortzet, ik wens hem hiermee veel succes;
- René du Bois en Kees Bos, die als derdejaarsstudenten nuttig werk voor me hebben verricht aan respectievelijk de Kelvin Probe en inversiealgoritmen;
- Bart Norbert, voor het samen bedenken van vele meetopstellingen en "6 mei tools". Verder Bram Huys, die deze dan vervolgens ging bouwen, snel en deskundig. Ook de vlotte hulp van Nico Schellenbach met diverse experimenten heb ik zeer gewaardeerd;
- Ruben Abellon van het IRI voor de transmissiespectra aan PET, Hans Lamers en Robert Jan van Wijk van Acordis voor de laserdefecten, Jan Veurink van Acordis en Ben Norder voor de gravimetrische wateropnamemetingen, Maarten ter Heerd voor de hulp met "zijn" Excimer laser en Piet Droppert voor de diëlektrische metingen;
- Verder ook alle andere medewerkers en studenten van Materiaalkunde voor de prettige tijd die ik met jullie heb gehad.

Tenslotte wil ik nog mijn ouders bedanken voor hun onvoorwaardelijke vertrouwen en aanmoedigingen. Dat laatste geldt in het bijzonder ook voor Cécile, die mij mijn gehele promotie met raad en daad gesteund heeft.