Application of Neural Network Analysis

to the Thickness-Shear Mode Chemical Sensor

by

Lan Nang Bui

A Thesis submitted in conformity with the requirements

for the degree of Doctor of Philosophy in the

Department of Chemistry

University of Toronto

1998

Canada

ABSTRACT

This thesis presents the application of artificial neural networks (ANN) to the self-referent calibration of thickness-shear mode (TSM) acoustic wave chemical sensors. Spectrum analysis of impedance measurements affords complete characterisation of the TSM sensor, which includes the use of the Butterworth-Van Dyke (BVD) equivalent circuit to quantify the electrical responses. The multi-parametric nature of this method and a novel weight-adjustment procedure, applied to ANN calculations, were utilised to effect a method of calibration in the presence of interferents. A network is trained, using exemplary I/O data acquired for a potassium chloride (KCl) system, to predict unknown outputs *i.e.* concentration, given four sets of measured inputs *i.e.* series resonant frequency (F_S), parallel resonant frequency (F_P), motional capacitance (C_m) and motional resistance (R_m). The trained and tested networks achieved high predictive efficiency with errors in the range of 2%-7%. An interferent, ethanethiol, is added to test the robustness of the trained network and was found to adversely affect the predictive ability of the network. The magnitudes of the weights, which are associated with the set of inputs deemed to be most affected by the interferent (F_S), are adjusted to minimise this deterioration. The resultant network, calibrated for the interferent, achieved a similar predictive efficiency for adulterated samples as that achieved by the original network for unadulterated samples. This calibration procedure was extended to a dual-interferent system in which two input variables are affected. A significant decline in predictive ability was observed for the unadjusted network, render it practically unusable. A two-point weight adjustment process was performed on the affected weights with considerable success but the predictive ability did not come close to the original level. The manipulation of two out of four input parameters may have over-extended the resources of the network but

may prove to be more amenable for a larger network.

A parallel project was effected in which a chemically-selective platform was immobilised on a TSM device following activation with bifunctional thiols. The optimal surface activation scheme was found to be the self assembly of a mixture of bifunctional thiols of different chain lengths. An amino-terminated 25-mer DNA single strand was successfully coupled to the distal carboxylic group via carbodiimide chemistry. The viability of the immobilised DNA receptor was demonstrated through a hybridisation experiment with a complementary DNA fragment. In addition, the multi-parametric nature of network analysis proved to be sensitive to the hybridisation kinetic and a subsequent, capacitive-based conformational change in the coupled strands.

ACKNOWLEDGEMENTS

First and foremost, I wish to express my gratitude to my esteemed supervisor, Professor Michael Thompson. The completion of this thesis was made possible with his guidance, encouragement, friendship and above all, his belief in my ability. The time spent under his tutelage was not only informative but also formative with respect to my maturation as a person. In addition to chemistry, I know now that my Tequila limit stands at a pitiful four shots. My fondest hope is that our association will continue to flourish and that one day we will be able to complete that 7 km race.

I am also very grateful to my colleagues in the lab, without whom the last few years would have been a lot less enjoyable. First on this list must be Dr. David Stone, who was and still is always available for helpful advice, assistance and discussions. I enjoyed very much our collaborations and our shared enjoyment of what surely must be some of the worst puns ever invented. I would also like to thank Dr. Steve Vigmond, Dr. Mengsu Yang, Dr. Mark McGovern, Dr. Biljana Cavic, Dr. Paul Li, Dr. Zhiping Deng, Dr. Larisa Cheran, Michelle Furtado, Anil Deisingh, Shakour Ghafouri, Zeynep Morel, Nardos Tassew, Emma-Louise Lyle and Blanca Granozi for all their helps, friendships and above all, for putting up with my somewhat sadistic sense of humour.

It is without a doubt that the greatest joy I have experienced during the last few years has been derived from my wife and son. It is their love that helped me through all obstacles and made me what I am today. I am thankful for their love, encouragement and infinite patience.

Finally, I want to thank my parents for their immeasurable love and support. Without their sacrifices for the last thirty years, none of my accomplishments would have been possible. This thesis is dedicated to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

A major problem since the inception of chemical and biosensor technology has been the quality of the signal-concentration relationship with respect to the time course of measurements. Lack of reproducibility in concentration values can be traced to changes in the performance characteristics of the concerned device over time and to perturbations caused by matrix interference. The incorporation of an artificial neural network (ANN's) into a sensor system may assist in solving this problem by providing a self-referent calibration mechanism for matrix effects. The main advantage of the neural network calculation over classical regression analyses is its ability to model, at an implicit level, any number of complex, non-linear data sets that would be impossible, or at least extremely difficult, to fit to an explicit mathematical model. This advantage has resulted in the application of neural network analyses to numerous fields, among them spectral interpretation [1-4], process control [5-7], prediction of protein structure [8-9] and various sensor systems. [10-17] More specifically ANN's have also been utilised in piezoelectric-based sensor systems but the focus has been primarily on arrays of single-output sensors, with the output usually being the series resonance frequency (F_S). [18-22] This single-output array architecture, in addition to generating extra complexity in theory and application, has served to limit the utility of neural networks to pattern recognition and classification. In the present work, spectrum analysis of the impedance measurements of a single sensor, using a network analyser, provides multidimensional information thus effectively overcoming these drawbacks. Such multi-dimensionality, in conjunction with a novel weight-adjustment procedure, may provide a calibration mechanism that can be used to eliminate a number of matrix effects.

1

TSM THEORY

The sensor platform central to this project is the thickness-shear-mode (TSM) acoustic wave sensor which in turn is based on an AT-cut piezoelectric quartz crystal transducer. The phenomenon known as piezoelectricity was discovered by the Curie brothers in 1880.[23] It is defined as *"...an electric polarisation produced by mechanical strain in crystals belonging to certain classes, the polarization being proportional to the strain and changing sign with it..."*.[24] Conversely, a polarisation within the crystal lattice, upon application of a voltage across the crystal, is accommodated by the displacement of atoms, leading to mechanical deformation in the crystal. This inverse piezoelectric effect, predicted by Hankel[25] and confirmed by the Curie brothers in 1881, is the basis for operation of the TSM.

A quartz crystal is essentially an electromechanical transducer that converts electrical energy to mechanical energy and *vice versa*. The electromechanical coupling and stresses resulting from an applied electric field depend on the crystal symmetry, the angle of cut relative to the crystallographic axes and the configuration of the excitation electrodes used to apply the electric field across the crystal. The various modes of electromechanical coupling result in different types of acoustic waves, modes of propagation and particle displacements. A comprehensive review of this topic was presented by Thompson and Stone.[26] A TSM device is prepared from a wafer, cut from a single quartz crystal at an angle of 35°15' relative to the z-axis[26] *i.e.* an AT cut, with metal electrodes deposited on opposite sides of the crystal (figure 1). Application of an alternating potential to the electrodes induces a displacement of the crystal lattice and produces a shear-mode bulk acoustic wave in the thickness direction, parallel to the electric field and perpendicular to the atomic displacements. This acoustic wave exists as a standing wave, with displacement maxima occurring at
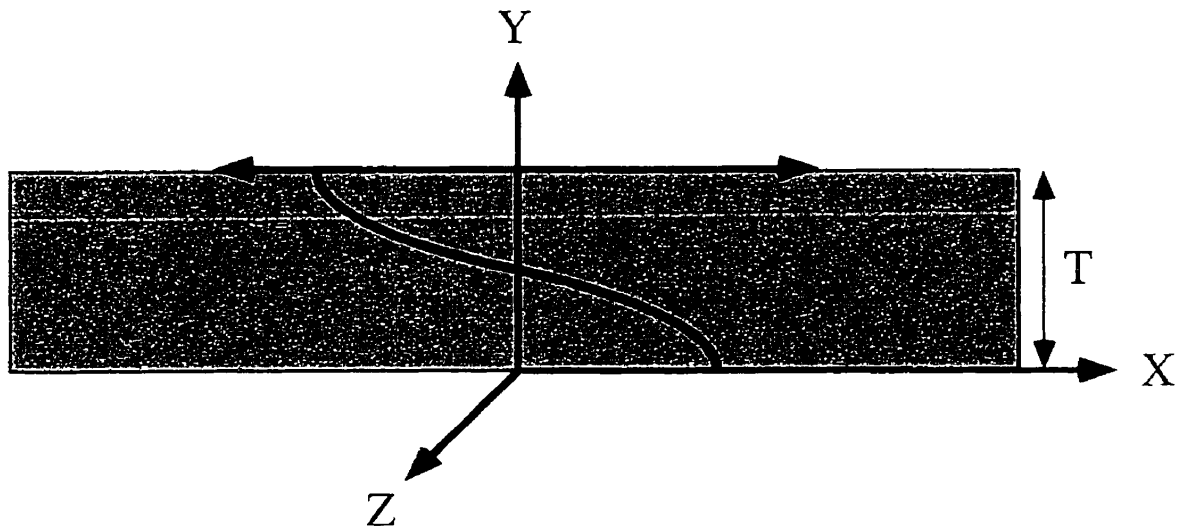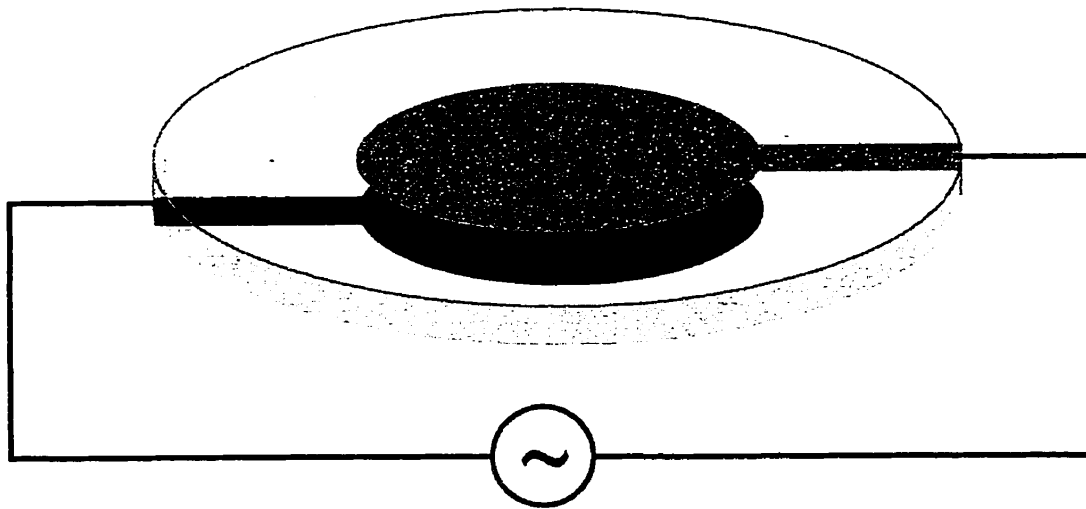
Figure 1: Thickness Shear Mode (TSM) device and wave propagation.

the crystal faces. The resonant frequency is determined by the physical characteristics of the quartz substrate and the thickness of the crystal since these parameters directly affect the wave velocity through the crystal. Since a rigidly adhered, thin coating on the electrodes can be treated as an extension of the quartz material and thus an increase in the thickness of the wafer, a shift in the resonance frequency can be correlated quantitatively with a change in mass on the surface of the crystal. This phenomenon was originally described by Sauerbrey, whose analysis was based on the increase in the resonant wavelength as a result of an increase in the effective thickness of the device.[27] An equation was derived for an AT-cut crystal, relating the proportional change in the resonant frequency to the deposited mass:

$$\Delta f = -\frac{2 f_q^2 \Delta m}{A \sqrt{\rho_q \mu_q}} \tag{1}$$

where Δf is the measured/expected frequency shift, f_q is the fundamental resonant frequency of the unloaded crystal, Δm is the total change in mass, A is the surface area of one face of the crystal in cm^2 with μ_q and ρ_q being the shear modulus and the density of the quartz crystal, respectively. For AT-cut crystal, $\mu_q = 2.947 \times 10^{11}$ g cm^{-1} and $\rho_q = 2.648$ g cm^{-3} which for typical operating frequencies and crystal sizes, predicts a detection limit on the order of 10^{-12} g [28]. Although this equation applies to experiments performed in vacuum or the gas phase, with infinitesimally thin films having the same acoustic impedance as that of the quartz crystal, it is not rigorously valid for thicker films. A more thorough approach by Miller and co-workers treated the device and the added mass to be a compound resonator [29]. It was demonstrated that for relatively small mass loading, a series expansion of the compound resonator

4

yields Sauerbrey's equation with higher-order terms. Lu and Lewis reduced this expansion to a simpler form, which is valid for frequency changes of up to 15% [30]. However, it was established that, for most materials, the Sauerbrey equation is valid for rigid overlayers up to a mass load limit of $\Delta m/m = 2\%$, where m is the mass per unit area of the unloaded crystal. [31]

TSM Operation in Liquid Phase

Prior to 1980, it was believed that damping viscous forces would preclude the liquid-phase operation of the TSM device. However, it was demonstrated in 1980 that the crystal could operate in contact with a liquid if one side of the device is kept exposed to a gas.[32] The behaviour of the TSM device when operated in liquids does not adhere to the Sauerbrey equation because significant losses of acoustic energy occur through transmission to the surrounding fluid. Energy from the crystal couples to the adjacent liquid, generating acoustic waves with decay lengths on the order of 1 μm. Bulk and interfacial properties, such as viscosity, density, conductivity, viscoelasticity, surface free energy, surface roughness, surface stress, molecular slippage, viscous energy losses and dielectric effects, have all been invoked to account for the change in the resonant frequency upon liquid loading and the various theories to date are briefly reviewed below and summarised in Table 1.

The first argument, based upon an empirical formulation, was presented by Nomura and Minemura.[32] The change in the resonant frequency, Δf, is ascribed to the specific density and conductivity of the adjoining solution. Nomura and Okuhara later showed that Δf, in an organic solvent containing no electrolyte, is influenced by the density and viscosity.[33] Through dimensional analysis, Bruckenstein and Shay [34] derived a model outlining the effects of viscosity and density on Δf, which can be applied to a crystal with one or two faces in contact with the liquid. A diffusion-momentum analogy was used by the authors to calculate the effective thickness of the layer of fluid deemed to be coupled to the surface. This was then correlated to the frequency shift through Sauerbrey's equation. A similar model, proposed by Kanazawa and Gordon in the same year [35], treats the quartz as totally lossless elastic solid and the liquid as a purely viscous fluid. Despite the different basis for the two

6

Equations	Comments	Authors
$\Delta f = A_1 k_L^{0.611} - B_1(\rho_L - 1)^{1.02}$	Empirical formulation. Conductivity and specific density.	Nomura and Minemura[32]
$\Delta f = A_2 \eta_L^{0.5} + B_1 \rho_L^{0.5} - C$	Empirical formulation. Viscosity and density.	Nomura and Okuhara[33]
$\Delta f = -2.26 \times 10^{-6} f_o^{3/2} (\eta_L \rho_L)^{1/2}$	Viscosity and density of bulk.	Bruckenstein and Shay[34]
$\Delta f = -f_o^{3/2} (\dfrac{\eta_L \rho_L}{\pi \mu_Q \rho_Q})^{1/2}$	Viscosity and density of bulk.	Kanazawa and Gordon[35]
$\Delta f = A_3 \Delta m_L = A_3 \dfrac{\rho_L \varepsilon}{2}$	Surface roughness.	Schumacher et al.[43]
$f_0 - f_m = A_4(P - P_m)^2$	Surface stress.	Haeusler et al.[38]
$\Delta f = -k_1 \Delta(\eta_L \rho_L)^{1/2} + k_2(\Delta \varepsilon_L)$	Hydrodynamic coupling analysis and liquid dielectric constant.	Hager[39]
$\Delta f = A_5 + A_6 \eta_L^{0.5} + A_7 \eta_L^{0.5} - A_8 \varepsilon_L$	Hydrodynamic coupling analysis and liquid dielectric constant	Yao and Zhou[42]

7

$$\frac{\Delta f}{f_0} = -(\frac{f_o \eta_L \rho_L}{\pi \eta_Q \rho_Q})^{1/2} \left\| \tanh(k_3 h) \right\|$$	Piezoelectric effects.	Shana and Josse.[40]
$$Z_f = \frac{N\pi}{4\kappa^2 \omega_s C_o}(\frac{G\rho_f}{C_{66}\rho_q})^{1/2} \tanh(j\omega T_f(\frac{\rho_f}{G})^{1/2})$$	Rheological properties.	Martin et al.[37]
$$\frac{\Delta f_s}{f_o} = \frac{\kappa^2}{\pi^2}\frac{\varepsilon_{22}}{\varepsilon_{22}+\varepsilon_L}\frac{\sigma^2}{\sigma^2+\omega^2(\varepsilon_{22}+\varepsilon_L)^2}$$	Acoustoelectric coupling.	Shana and Josse.[40]
$$R_m = \frac{A}{K^2}(2\pi f_o \eta_L \rho_L)^{1/2}$$	Equivalent circuit.	Muramatsu and et al.[41]
$$R_m = \frac{\eta_Q}{\mu_Q C_1}(\frac{\omega}{\omega_s})^2 + \omega_s\frac{L_1}{N\pi}(\frac{2\omega\rho\eta_L}{\mu_Q\rho_Q})^{1/}$$	Equivalent circuit.	Martin and et al.[44]

Table 1: TSM Operation in Liquid.

models, the relationship between the frequency shift and the properties of the quartz and bulk liquid is the same as that predicted by Bruckenstein and Shay. The change in the resonant frequency was deemed to have arisen from the coupling of the crystal oscillation with the liquid, resulting in a standing shear wave with a damped, propagating component into the liquid (figure 2). The decay length of this component varies as the square root of the viscosity of the bulk liquid and constitutes the effective thickness of the liquid treated as a rigid layer. Both models assume non-slip conditions, *i.e.* the transverse displacement of the surface of the device is identical to that of the adjacent liquid layer, and they are most rigorous only when the contacting liquid is perfectly insulating and the electrode surface is ideally smooth. In particular, these models predict that the change in series resonant frequency is proportional to the acoustic impedance of the liquid. Specifically, it was concluded that only a thin layer of liquid at the surface of the quartz will undergo displacement and thus the response of the device will be a function of the characteristics of this layer. These two models appear to form a bridge between the classical Sauerbrey gas-phase mass response and the properties of a thin boundary film of liquid. However, by considering only the mechanical resonant frequencies and ignoring the electrical properties of the TSM device, these models do not address the operation of the sensor on a fundamental basis and can predict only some aspects of its behaviour in liquid.

A comprehensive two-layer model was proposed by Reed *et al.*, based on the same non-slip boundary conditions.[36] The electrical admittance of a crystal, with one side exposed to a viscoelastic medium, was described in terms of the physical properties of the system. The description was based on the detailed matrix equations for piezoelectric quartz and includes the piezoelectric effect of the quartz transducer and the shear modulus and viscosity of the layer. The boundary conditions included
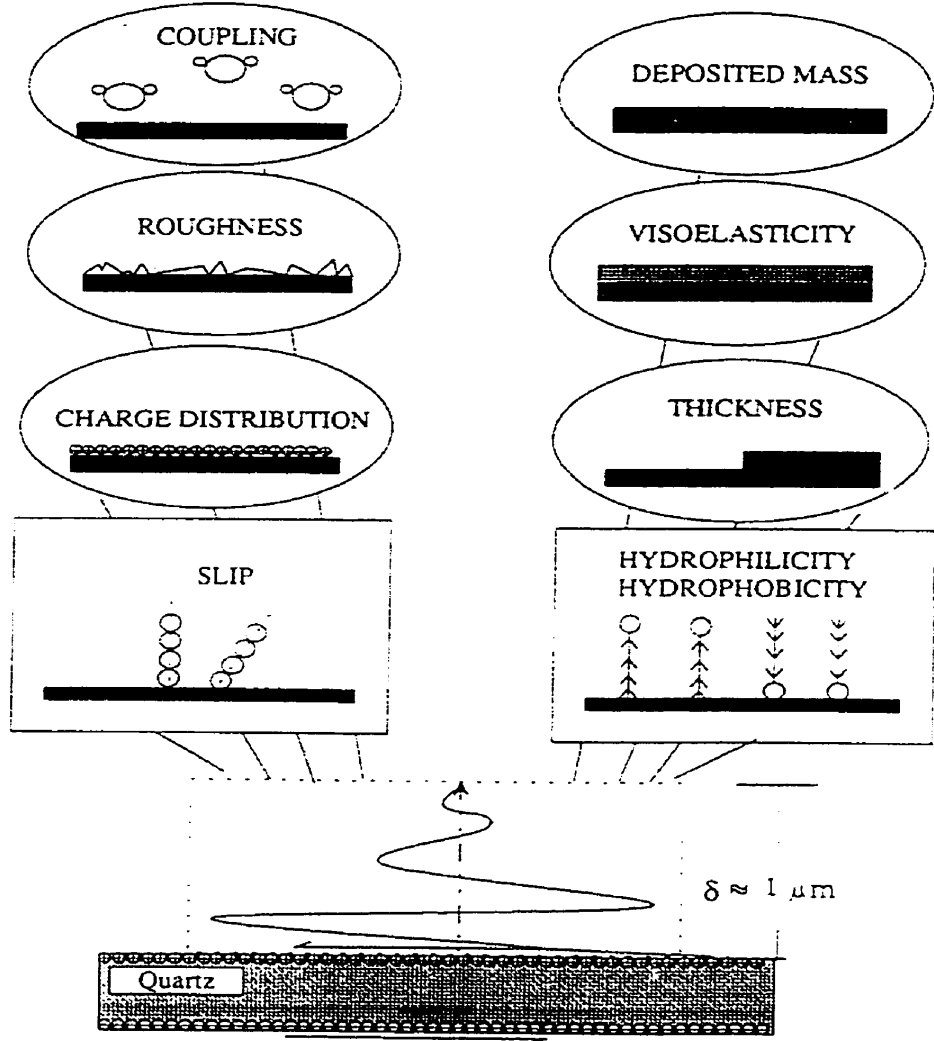
Figure 2: Factors affecting the responses of the TSM device.

no slip, no stress at the free surfaces and the application of a driving voltage to both sides of the quartz crystal. The viscous component in the Hook's Law equation was included to account for the energy dissipation in the crystal. When the adjacent liquid layer is considered to be infinitely thick and purely viscous, this model mirrors that of Kanazawa and Gordon.[35] Using this model, Martin and co-workers derived an equivalent circuit model for the electrical behaviour of the TSM device in liquid.[37] In their work, the viscous film is replaced by an equation for liquid loading derived from the Navier-Stokes equation. The deposited mass was considered to be an infinitesimally thin film and the no-slip condition was invoked. This model allowed both mass and viscous loading to be measured simultaneously. Haeusler and co-workers considered the influence of surface stress on the resonant frequency and proposed a theory in which a parabolic dependence of the resonant frequency on the hydrostatic pressure at the surface of the crystal could be related to the energy stored in the quartz.[38] Hager derived a model in which viscous energy losses, fluid velocity at the surface of the crystal and the dielectric effect of the liquid were considered.[39] Josse and co-workers derived an equation for the observed frequency shift of a TSM device, with one side immersed in a dilute conductive solution, by applying the appropriate electrical and mechanical boundary conditions at the interface and solving the characteristic equations for each medium [40]. Muramatsu and co-workers have also demonstrated a llinear relationship between Δf and $(\rho_L \eta_L)^{1/2}$ by considering the resistance of an equivalent circuit for alcohol-water solutions [41]. This correlation deviates from linearity when the crystal is exposed to highly viscous liquids or when both sides of the device are in contact with water. Furthermore, according to the work of Yao and Zhou, the frequency response of a TSM device in liquid depends on the dielectric and conductance effects of the liquid [42]. Their model resembles that which

11

was derived by Nomura and Okuhara. [33]

These early theories all lack consideration for the boundary conditions that exist at the crystal - liquid interface *i.e.* surface roughness, interfacial viscosity, surface free energy and interfacial slippage (figure 2). With respect to the first condition, when the surface of the resonator is rough, the liquid motion generated by the vibrating surface becomes more complicated. A variety of additional interaction mechanisms may arise from the generation of turbulent flow or compressional waves by protruding portions of the surface and from the trapping of liquid in pits and crevices. Schumacher and co-workers proposed that anomalously large shifts in Δf could be attributed to surface roughness.[43] The authors proposed a model of a corrugated surface comprising of hemi-cylinders with diameters ε. The entrapped liquid, which can be represented as a rigidly attached liquid layer having a thickness of $\varepsilon/2$, was deemed to be directly culpable for the additional change in the resonant frequency. A later study of surface roughness by Martin *et al.* led to the conclusion that this layer of trapped liquid is adequate explanation for any anomaly in the resonant frequency and there is no basis for either liquid ordering, or slippage, at the quartz-liquid interface.[44] The authors argued that for roughness features of dimension much less than the liquid decay length, the surface may be considered to be hydrodynamically smooth and the frequency response should depend only on the density-viscosity product. The wettability of a smooth surface is also considered to be unimportant with respect to its ability to couple to the liquid. As the features on the surface become comparable to or larger than the decay length, liquid trapping and compressional wave generation may become important mechanisms in the response of the device. It was also concluded that the extent of liquid trapping by a rough surface diminishes with increasing hydrophobicity. This effect was deemed to be wholly responsible for phenomena

previously attributed to slippage at the solid-liquid interface.

A thorough investigation of rough *vs.* smooth surfaces by Yang and Thompson concluded that significant contributions to frequency shifts can also arise from surface stress as well as shear and non-shear coupling phenomena.[45] Subsequent studies by Thompson *et al.*[46-49] have demonstrated that the response of the TSM device is controlled by the molecular boundary conditions that exist at the quartz - liquid interface. It was concluded that perturbation of acoustic energy propagation could be caused by a partial slip boundary condition at the interface. This tenet is central to the concept of interfacial viscosity at a solid-liquid junction, with one component in motion. The culmination of these investigations is a four-layer model, proposed by Thompson and Duncan-Hewitt, which applies the rigorous linear theory for piezoelectricity to describe the interfacial viscosity in term of an activation barrier to interfacial flow.[50] Using this model and the molecular theory of interfacial viscosity proposed by Krausz *et al.* [51], they were able to show that the response of the TSM device could be associated experimentally with the surface free energy, as determined by contact angle measurements. This, in turn, can be correlated to interfacial slip characteristics. Duncan-Hewitt and co-workers were able to demonstrate also the role of interfacial parameters in the determination of interfacial viscosity. It has been proven that the properties of a liquid adjacent to a solid interface are different from those of the bulk and these properties may indeed affect the behaviour of the entire system.[52-54] Figure 3 depicts the four-layer model of the TSM sensor with the first layer consisting of the quartz-electrode system. The authors made the assumption that that the electrodes on the quartz surfaces are thin and elastic enough to approximate a layer of quartz. The second layer is an ordered surface-adjacent layer of liquid, possessing greater density and viscosity than the bulk liquid, the thickness of which is a function of the solid surface-liquid interaction. The viscosity and density of this

13
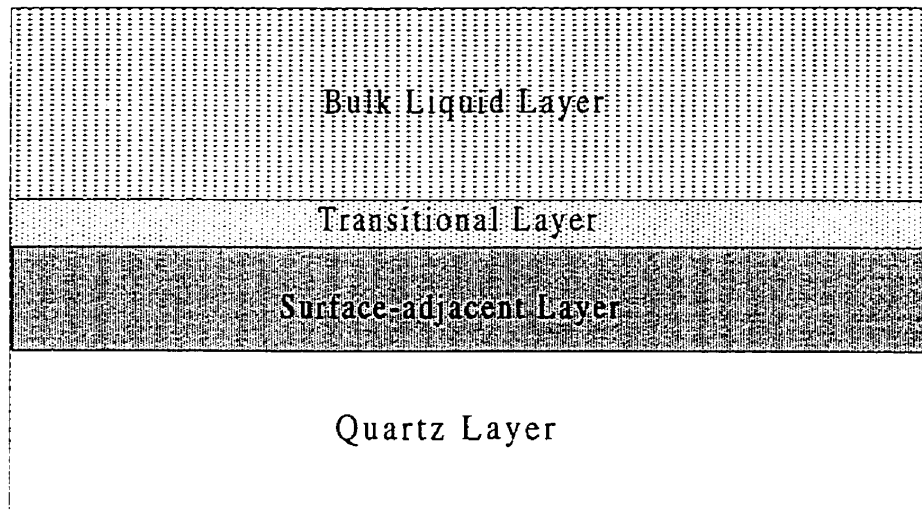
HYDROPHILIC SURFACE
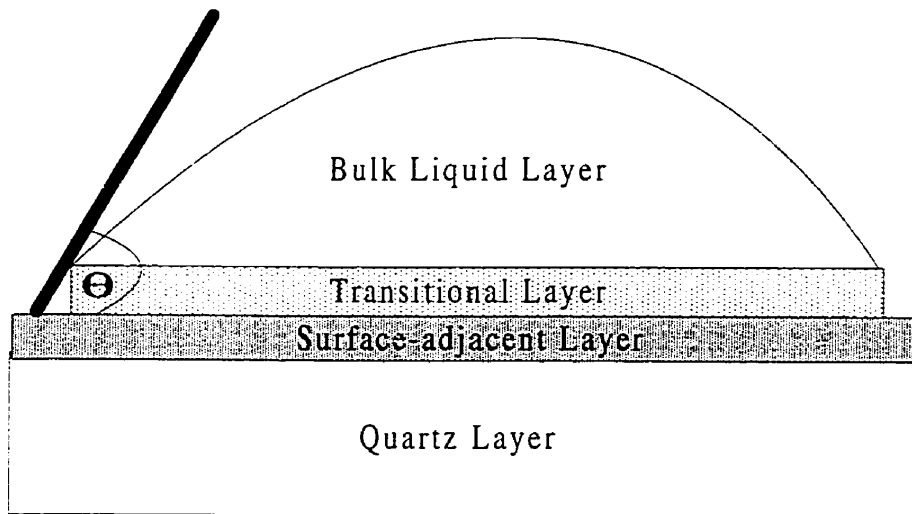
HYDROPHOBIC SURFACE

Figure 3: Four-layer model.

region are deemed to be five times that of the bulk. Churaev and co-workers have demonstrated that the apparent viscosity of water in a quartz capillary 0.1 μm wide is 1.5 times that of bulk water.[55] The thickness of the ordered surface-adjacent layer is correlated with experimental contact-angle data. The third layer is a thin transition region between the surface-adjacent layer and the bulk liquid. Its composition and behavior is influenced by the hydrophobicity of the system and can be predicted from contact-angle measurements. The thickness of this transition region is estimated, from theoretical and experimental considerations, to range anywhere from one to five molecular diameters. The density of this layer should approach that of the bulk liquid if the surface is completely wetted and that of the vapor phase for a hydrophobic surface. The final region directly affected by the device is the bulk liquid, with an approximately thickness of 3μm. All the liquid layers were deemed to be purely viscous and the no-slip condition was invoked for all the liquid-liquid interfaces. The solutions of the differential equations require that each layer be homogenous as well as the assumption of abrupt transitions between layers, but in reality the transitions across each layer are more gradual. This model correlates the viscosity and density of the interfacial layers with interfacial energetics, via the contact angle, and the predicted results are demonstrated to be quantitatively similar to those observed experimentally.

The authors also considered the validity of the no-slip boundary conditions, prevalent in earlier models. The possibility of slippage at a solid - liquid interface was previously proposed by Tolstoi [56] and Blake [57], demonstrated through surface force microscopy on thin films by Israelachvili [58] and later applied to describe the behaviour of the TSM device by Krim *et al.* [59] A similar conclusion was reached recently by Hayward in the derivation of a model relating equivalent circuit elements

to the dissipation of acoustic energy.[60] The first detailed study of molecular slip at the interface of an AT-cut TSM sensor was reported by Ferrante *et al.* [61] A model is derived through impedance analyses of hydrophobic and hydrophilic surfaces, in contact with water-glycerol solutions of various concentrations. A new quantity, the interfacial slip parameter α, is introduced as one of the boundary conditions for a two-layer model of the TSM sensor in liquid. The two-layer model predicates that the structure of the liquid in contact with sensor surface is not affected by the surface. This slip parameter is a complex-valued quantity and it is defined as the displacement of a particle of liquid in contact with the surface divided by the displacement of a particle on the surface of the sensor. The experimental impedance values of the TSM device, as measured by the network analysis method, were fitted to the theoretical expression by non-linear, iterative regression to find the slip parameter and several other crystal-related parameters. A mechanical model of molecular slippage is devised to explain the variation of the slip quantity with the viscosity of the liquid. It was concluded that if the forces between the liquid particles and solid particles were very strong, their displacements would be the same. In this case, α would be equal to 1 and there would be no slip at the interface. However, if the force between the liquid and solid was zero, then there would be no coupling between the two phases. In that scenario, movement by the solid particles on the surface would not induce the displacement of the liquid particles, thus α would be equal to 0. In most systems, the magnitude of α will not be at the extremes but will lie between 0 and 1. It was observed that there was some slip in liquids with low viscosity near that of water but very little slip in liquids of high viscosity near that of glycerol.

Recently, Hayward and Thompson extended the concept of interfacial slip parameter to a modified four-layer model.[62] A real coated sensor is modelled, with

the four layers being the quartz crystal, the metal electrodes, a polymeric sensing layer and an infinitely thick fluid medium. All four layers are assumed to have both elastic and viscoelastic properties. The bottom face of the crystal is considered to be a free surface, having an ideal electrode with neither mass nor thickness. Nine boundary conditions were invoked with the first three being the interfacial slip parameters between successive layers. These parameters were assumed to be complex-valued to allow phase shift between the displacements. This assumption yielded three equations. At the same three interfaces the shear stress must also be continuous, yielding a second set of three equations. The no-stress condition at the bottom face of the crystal gave rise to another equation. The other two boundary equations are the electrical forcing functions that cause the crystal to oscillate. This nine-equation model was used to fit network analysis data, obtained from water-glycerol solutions of various concentrations, with good agreement. With appropriate assumptions, correlations were also obtained with mass loading models by Sauerbrey [27] and Lu et al.[30], with non-slip liquid loading model by Kanazawa and Gordon [35] and with slip-invoked liquid loading model by Ferrante et al. [61]. The authors also adapted features from three earlier models to describe the behavior of the liquid at the surface of the quartz crystal. The fluid was considered to be viscoelastic, to be ordered at the interface and to slip at the interface according to Reed et al. [36], Duncan-Hewitt et al. [50] and Ferrante et al. [61], respectively. The series resonant frequency was shown to have a proportional relationship with the stiffness of the fluid medium and an inverse relationship with its viscosity. The two parameters were combined to form a complex shear viscosity that was interchangeable but not physically equivalent to the interfacial slip parameter. According to the authors, both viscoelasticity and interfacial slippage are important with the former being significant for more viscous liquids and the latter more dominant for thinner fluids.

17

The responses of the TSM device to a conducting fluid have also been well-documented. In a conducting medium, the acoustic wave generated by the crystal will interact with the ions and dipoles of the solution via viscous coupling and acoustoelectric coupling. Acoustoelectric coupling arises when the evanescent wave, generated by the propagating acoustic wave, interacts with the charge carriers and dipoles in the surrounding medium. The energy expended through this interaction would affect the response of the TSM device. The performance of the device is thus influenced by the electrical properties of the solution. Nomura et al.[32] and Yao et al.[42] have proposed empirical models, describing the relationship between the resonant frequency and the specific conductivity and the dielectric constant of the surround fluid, respectively (Table 1). Through the application of the appropriate electrical and mechanical boundary conditions at the interface and by solving the resulting equations, Josse and co-workers [40] derived an equation describing the frequency shift of a TSM device with one side immersed in a dilute conductive solution (Table 1). The change in F_S was described in terms of the dielectric constants for both quartz and liquid as well as the conductivity of the solution.

Other factors have been proposed which may affect the responses of the sensor: the formation of an electrical double layer at the interface and the fringing of the electrical field, created by a small spillover of the electrical field beyond the area bounded by the electrodes. According to Adamson [63], while it is present, the thickness of the double layer is much less than the decay length of the electric field (0.01 μm vs. 0.2 μm) and thus can be regarded as an added capacitance to the TSM sensor. Rodahl and co-workers suggested that acoustically induced surface charges could couple to charged species in a conducting liquid [64] but the inference of the existence of surface charges, or other microscopic phenomena such as electrical

18

double layers, was not necessary to account for the influence of the electrical properties on the resonant frequency. The contacting conductive liquid effectively enlarges the electrode area and changes the parallel resonant frequency. This has the implication that the fringing fields and the two electrode contacts may play an important role in both the resonant frequency and the energy dissipation response of a TSM device.

The authors also considered the changes in the stray capacitive and the resistive pathways caused by the liquid via its dielectric and conducting properties, respectively. The origin of a shunt capacitance, C_o^{eff}, in air can be attributed to the capacitor formed by the two electrodes sandwiching the quartz crystal. It was observed that the fringing field would cause a stray capacitance that contributed to the shunt capacitance of the TSM. The size of this stray capacitance depended on the geometry and the dielectric constant of the media that the field lines passed through; the higher the dielectric constant, the larger the capacitance. A significant increase in the capacitance would have a definite effect on the response of the TSM sensor.

NETWORK ANALYSIS

Prior to 1990, characterisation of the TSM device has been done exclusively through the oscillator method.[34,65] A common oscillator architecture would comprise of two transistor-transistor-logic (TTL) inverters connected in series to give a non-inverting amplifier *i.e.* there is zero phase shift between the input and the output voltage. The quartz crystal is connected from the output to the input of the amplifier to produce a positive feedback. Two criteria must be fulfilled for oscillation to occur: (a) the loop gain of the circuit must be unity, *i.e.* the feedback voltage is equal to the input voltage and (b) the phase shift must be zero. The configuration of the electrodes on the crystal gives rise to an electrical capacitance and thus dual resonance. The two frequencies at which the phase shift is zero are the series resonant frequency, F_S, and the parallel resonant frequency, F_P. However, for any particular oscillator setup, the unity loop gain criterion can be satisfied for only one resonant frequency. In most cases, the circuit is designed to oscillate at F_S although the inverse configuration is possible.

The oscillator method suffers from three glaring limitations. Since it yields only one resonant frequency, the crystal is only partially characterised. Moreover, this frequency is not wholly characteristic of the device but can be modulated by simply changing the value of a capacitor in series with the crystal. Also, the circuit will cease to oscillate if the crystal is immersed in liquid or if it is exposed to a solution of high viscosity. This fundamental deficiency exists because at high viscosity, the phase shift of the crystal is less than zero. Thus, the second criterion for oscillation is not satisfied. Attempts to overcome the first limitation included the measurement of the output voltage [39], the feedback voltage of the oscillator amplifier [48] and the use of an

impedance analyzer to measure the motional resistance (R_m) [66], in addition to the resonant frequency. However, two quantities do not describe satisfactorily the behaviour of the TSM device thus the characterisation is still incomplete. The concept of impedance analysis was extended by Thompson et al. through the use of a network analyzer.[46] This is a passive method in which sinusoidal voltages, on and reflected from the quartz crystal, and the current flowing through it are measured for 401 frequency points about resonance. The magnitude and phase of the impedance are calculated at each frequency from the voltage and current, the plots of which are illustrated in figure 4. From these plots, up to eight parameters can be determined directly and several more electrical quantities can be derived from equivalent circuit analysis of the impedance curves. This technique allows for the complete characterization of the quartz sensor.

BVD EQUIVALENT CIRCUIT

The theory concerning the electrical behaviour of a piezoelectric crystal has been derived by Cady [67] and summarised by Bottom [68]. The equation of motion for each point of the quartz plate was considered and the result of the analysis was an expression for the current density conducted by the crystal, in terms of the voltage across the crystal. This complex-valued quantity, the admittance, is a function of the frequency and of the properties of the crystal. The reciprocal of the admittance, the impedance, can be interpreted as an equivalent circuit after some mathematical manipulations and approximations. An equivalent circuit is a combination of ideal electrical components, which responds to an applied voltage or current in the same way as the device itself. This electrical model will describe exactly the impedance properties of the device, as a function of frequency. The Butterworth-Van Dyke (BVD) equivalent circuit (figure 5) corresponds to the thickness-shear mode of
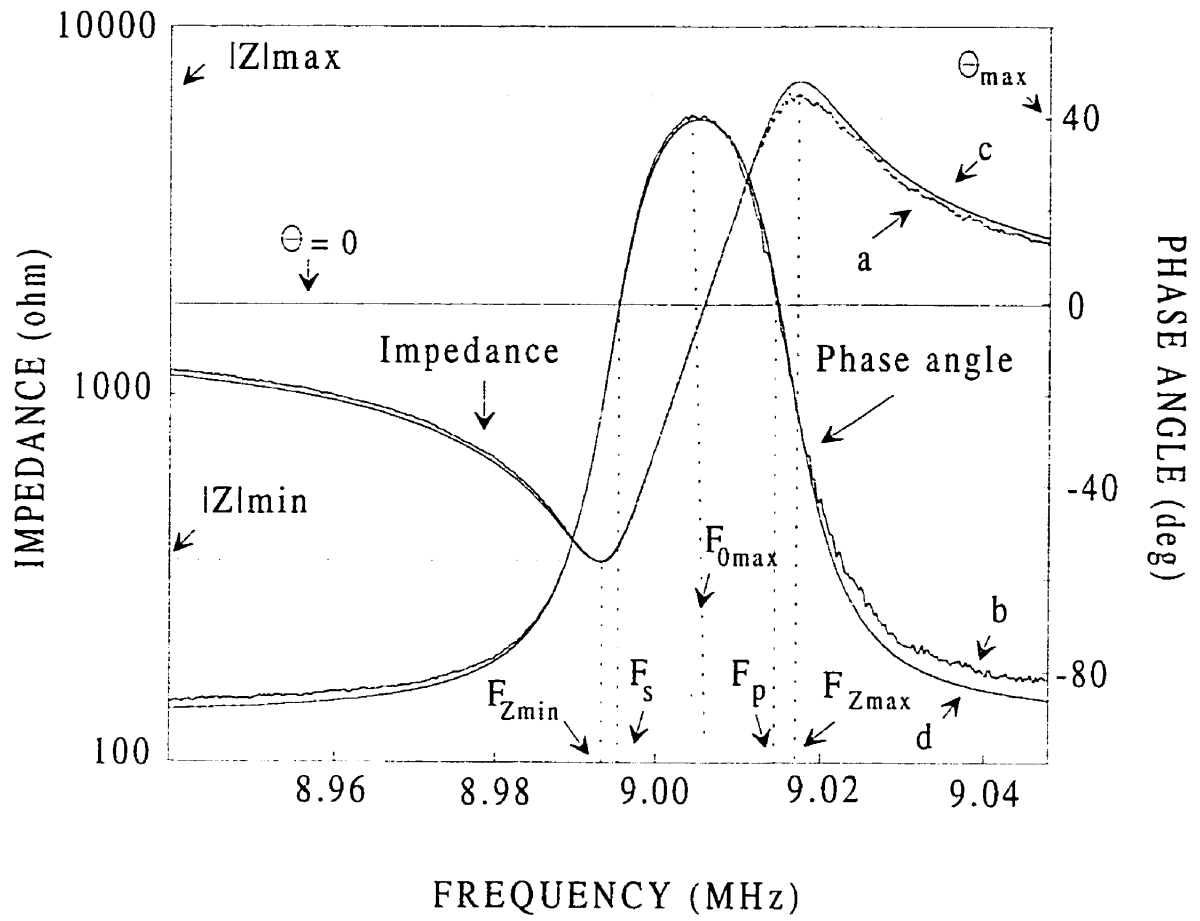
Figure 4: Impedance (a) and phase (b) traces derived from network analysis. (c) and
(d) are the fitted traces using the electrical elements.

C_0

R_m L_m C_m

EQUIVALENT CIRCUIT PARAMETERS

$$\frac{-j}{\omega C_0}$$

R_m $j\,\omega L_m$ $\dfrac{-j}{\omega C_m}$
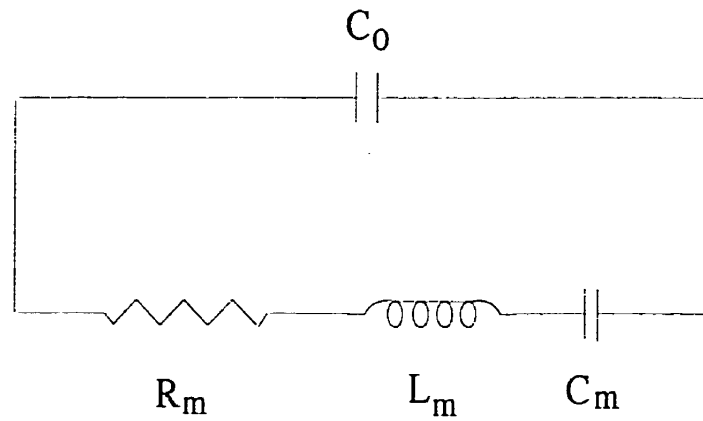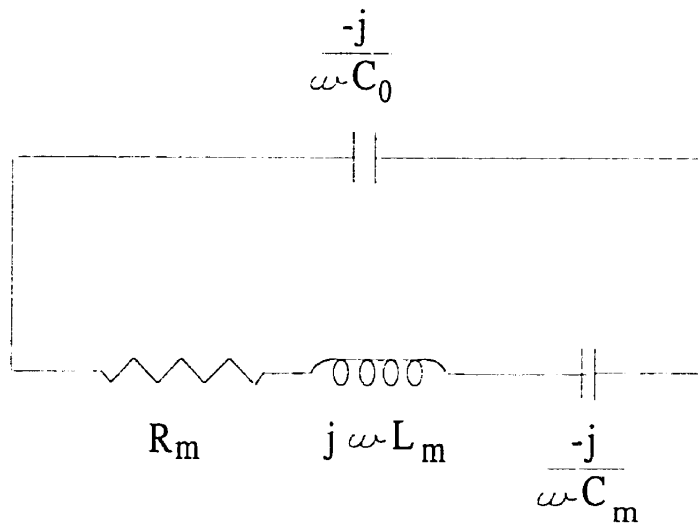
IMPEDANCES OF E. C. PARAMETERS

Figure 5: Equivalent circuit parameters.

vibration of the bulk acoustic wave device. It comprises of a capacitor (C_o) in parallel with a series combination of a resistor (R_m), an inductor (L_m) and a capacitor (C_m). The electrostatic capacitor (C_o) accounts for the capacitance of the two electrodes sandwiching the insulating quartz. Electro-mechanical coupling gives rise to the additional motional contribution (R_m, L_m, and Cm) in parallel with the electrostatic capacitance. These are associated with the vibrational motion caused by the piezoelectric effect in the resonant region. Thus the electrostatic element will dominate the impedance away from resonance while the motional arm assumes control near resonance. All four circuit elements can be expressed as a function of the physical characteristics of the quartz plate. These are listed below along with their respective impedances.

$$R_m = \frac{e^3 \sigma_d}{8A\varepsilon^2} \qquad \text{imp.} = R_m \qquad (2)$$

$$L_m = \frac{e^3 \rho}{8A\varepsilon^2} \qquad \text{imp.} = j\omega Lm \qquad (3)$$

$$C_m = \frac{8A\varepsilon^2}{\pi^2 e\mu_Q} \qquad \text{imp.} = -\frac{1}{j\omega C_m} \qquad (4)$$

$$C_o = \frac{k\varepsilon_0 A}{e} \qquad \text{imp.} = -\frac{1}{j\omega C_0} \qquad (5)$$

where e is the thickness, σ_d is the dissipation coefficient, A is the area, ϵ is the piezoelectric stress constant, ρ is the density and μ_Q is the elastic constant of the quartz crystal. The dielectric constant of quartz, k, and the permittivity of free space, ϵ_o, are unique to the equation describing the electrostatic capacitance, which, by not containing ϵ, is independent of the piezoelectric effect.

24

Each individual component can be correlated to the physical properties of the device as well as those at the surface and of the surrounding medium. In practice, the impedance and phase curves of a TSM device are measured experimentally and the electrical elements are fitted to the model to extract information concerning the solid-liquid interactions. The motional resistance, R_m, represents the dissipation of electrical energy in the TSM device, both internally and externally. Internally losses involve conversion into thermal energy due to internal friction while externally losses include mechanical losses in the mounting system and dissipation into the surrounding medium via the decaying acoustic waves. While the internal losses depend only on the physical properties of the quartz plate, the external dissipation is a function of both the bulk properties of the fluid and the nature of the solid-liquid interaction at the interface. L_m, the motional inductance, is the inertial component related to the displaced mass during oscillation. It represents the mass of the quartz plate, the coating on its surface and the effect of the contacting fluid. The motional capacitance, C_m, is the compliance of the quartz element, representing the energy stored during oscillation. The electrostatic capacitance, C_o, is simply the capacitance of the two parallel electrodes on both faces of the TSM device.

As can be seen from figure 4, there are several characteristic frequencies that can be extracted from the experimental impedance and phase angle curves. These frequencies have been derived theoretically in terms of the circuit elements to demonstrate their relationship.[67-68] The impedance of each of the elements can be expressed in term of the angular frequency, ω (in rad/s), where $\omega = 2\pi f$ and f is the frequency of the voltage source (figure 5). Z, the complex-valued impedance of the BVD circuit, can be written as the impedance combination of the four individual components.

$$Z = Z_m * Z_o / (Z_m + Z_o) \qquad (6)$$

where

$$Z_m = R + jX_m \qquad\qquad Z_o = jX_o \qquad (7)$$

and

$$X_m = \omega L_m - (1/\omega C_m) \qquad\qquad X_o = -(1/\omega C_o) \qquad (8)$$

or in standard form,

$$Z = R + jX \qquad (9)$$

where R is the real part of Z, the resistance, and X is the imaginary part, the reactance. Both R and X are functions of R_m, L_m, C_m, C_o and the angular frequency, ω. By definition, the admittance, Y, is the reciprocal of the impedance Z.

$$Y = 1/Z = G + jB \qquad (10)$$

where G is the conductance and B is the susceptance. These quantities can be expressed in term of the circuit elements.[67,68]

$$G = \frac{R_m}{R_m^2 + \left(\omega L_m - \dfrac{1}{\omega C_m}\right)^2} \qquad (11)$$

$$B = \frac{-\left(\omega L_m - \dfrac{1}{\omega C_m}\right)}{R_m^2 + \left(\omega L_m - \dfrac{1}{\omega C_m}\right)^2} + \omega C_0 \tag{12}$$

The magnitude (|Z|) and the phase angle (θ) of Z, in radians, are given below:

$$|Z| = \sqrt{R^2 + X^2} \tag{13}$$

$$\theta = \tan^{-1}\left(\frac{X}{R}\right) \tag{14}$$

Physically, Z is the voltage across the crystal divided by the current flowing through it. The phase angle, θ, is a measure of the time difference between the voltage maximum and the current maximum. If θ is positive, the voltage leads the current and if θ is negative, the voltage lags the current. Since from basic complex number relations one can write

$$|Z| = |Y|^{-1} \qquad \text{and} \qquad \arg Z^{-1} = -\arg Y \tag{15}$$

$$\therefore \qquad |Z| = \sqrt{R^2 + X^2} = \left(\sqrt{G^2 + B^2}\right)^{-1} \tag{16}$$

and

27

$$\tan\theta = \frac{X}{R} = -\frac{B}{G} \tag{17}$$

substituting $X = -R\dfrac{B}{G}$ into equation 16 gives

$$\left(R^2 + X^2\right) = \left(R^2 + R^2\frac{B^2}{G^2}\right) = \frac{1}{G^2 + B^2} \tag{18}$$

which reduces to

$$R = \frac{G}{G^2 + B^2} \tag{19}$$

and similarly substituting $R = -X\dfrac{G}{B}$ into equation 16 and doing the appropriate

reduction yields

$$X = \frac{-B}{G^2 + B^2} \tag{20}$$

The condition for a circuit to be resonant is the imaginary part of Z should be equal to zero, that is Z is a pure resistance. This is satisfied at zero phase since for $\tan\theta = 0$, $B/G = 0$. Therefore, by dividing (B) by (G) and equating to zero, one obtains

28

$$\frac{1}{R_m}\left[-\left(\omega L_m - \frac{1}{\omega C_m}\right)+\left\{R_m^2 +\left(\omega L_m - \frac{1}{\omega C_m}\right)^2\right\}\omega C_0\right]=0 \qquad (21)$$

$$\left\{R_m^2 +\left(\omega L_m - \frac{1}{\omega C_m}\right)^2\right\}\omega C_0 = \omega L_m - \frac{1}{\omega C_m} \qquad (22)$$

$$\frac{R_m^2 \omega C_0}{\omega L_m - \frac{1}{\omega C_m}}+\omega C_0\left(\omega L_m - \frac{1}{\omega C_m}\right)-1=0 \qquad (23)$$

The motional resistance, R_m, represents the energy dissipation of the device but in the gas phase, $R_m \approx 0$. By assuming $R_m = 0$ and by solving the resulting quadratic equation, one obtains

$$F_R = \frac{1}{2\pi}\sqrt{\frac{1}{L_m C_m}} \qquad (24)$$

$$F_A = \frac{1}{2\pi}\sqrt{\frac{1}{L_m C_m}+\frac{1}{L_m C_0}} \qquad (25)$$

where F_R is the resonant frequency and F_A is the anti-resonant frequency. For cases in which R_m is not negligible, i.e. upon liquid loading, the solutions must include the effect of the motional resistance to obtain

$$F_S = F_R\left(1+\frac{r}{2Q^2}\right) \qquad (26)$$

29

$$F_P = F_R \left(1 + \frac{1}{2r} - \frac{r}{2Q^2} \right) \tag{27}$$

where $r = C_0 / C_m$ and $Q = \omega L_m / R_m$. Q, termed the quality factor, is defined as the ratio between the energy stored and the energy dissipated per resonant cycle. The quality factor is unique to each crystal and determines its ability to control the frequency. F_S and F_P are the zero-phase frequencies at which Z is a pure resistance. The low frequency of zero phase, F_S, is termed the series resonant frequency and the high frequency of zero phase, F_P, is called the parallel resonant frequency. They are the same as those that can be measured separately using the oscillator method. Network analysis allows for the simultaneous determination of both frequencies, along with other parameters. Two more characteristic frequencies, obtainable from the impedance curves, can be derived theoretically via circuit analysis. The frequencies of minimum and maximum impedances can be obtained by computing the modulus of the admittance Y and differentiating the resulting expression with respect to X.

$$F_{ZMIN} = F_R \left(1 - \frac{r}{2Q^2} \right) \tag{28}$$

$$F_{ZMAX} = F_R \left(1 + \frac{1}{2r} - \frac{r}{2Q^2} \right) \tag{29}$$

The admittance of a circuit can be depicted graphically to illustrate the origins of the characteristic frequencies and their relationships. The admittance diagram is a complex plane with the conductance (G) as the real axis and the susceptance (B) as the imaginary axis (figure 6). The admittance circle depicts the frequency, increasing
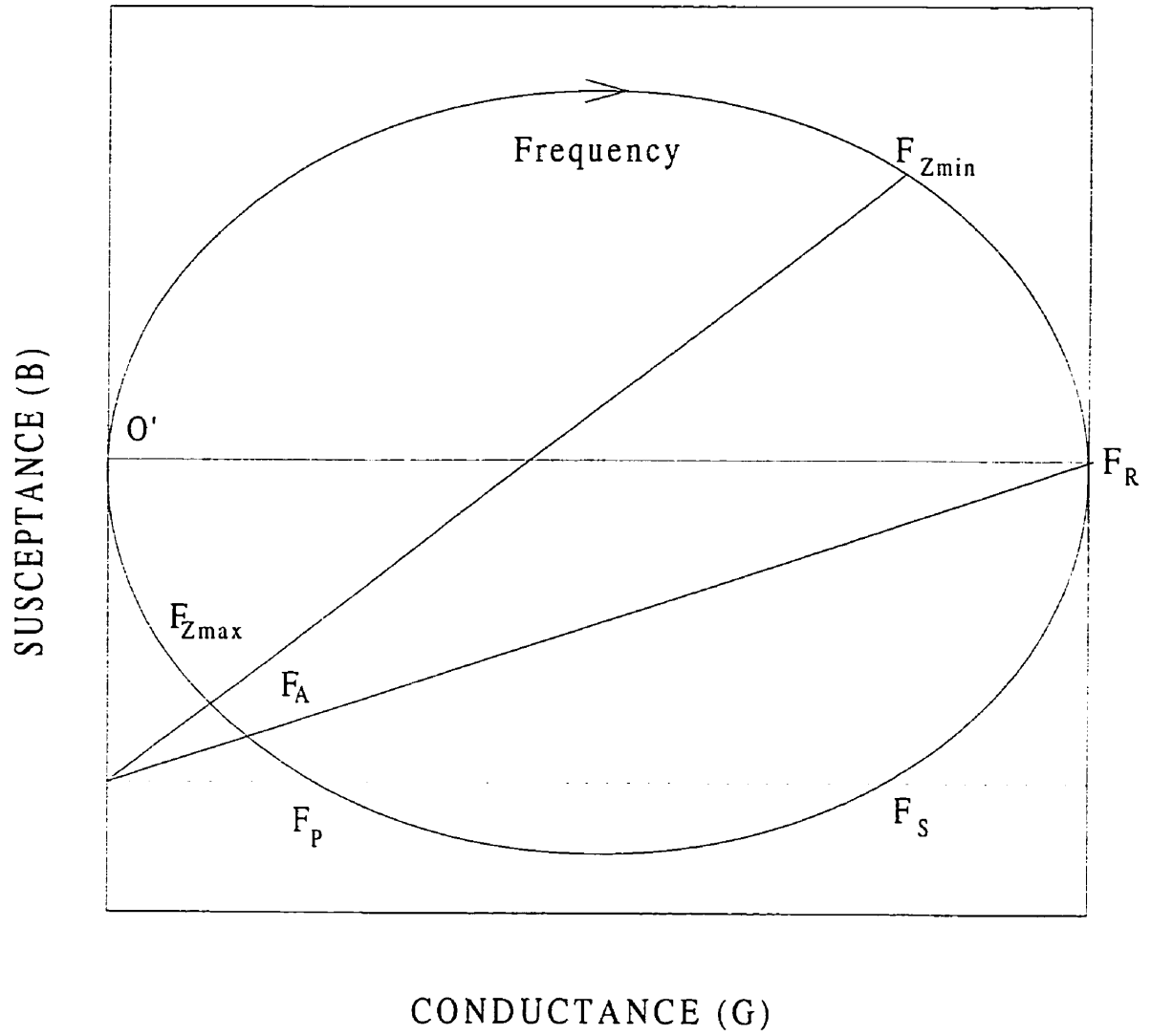
Figure 6: Admittance diagram.

in a clockwise direction, with any point on the circle representing the vector sum of G and B. The relative positions of all the characteristic frequencies are depicted on the frequency curve, in the following order: F_{ZMIN}, F_R, F_S, F_P, F_A and F_{ZMAX}.

In an actual experiment, F_S, F_P, F_{ZMIN}, F_{ZMAX} are measured directly from the impedance curves (figure 4), the circuit elements are derived from a least-squares fit of the curves while F_R and F_A can be determined by substituting the respective circuit parameters into equations 24 and 25. Other useful parameters that can be obtained from the magnitude and phase curves are the minimum impedance, Z_{MIN}, the maximum impedance, Z_{MAX}, the maximum phase, θ_{MAX}, and the frequency of maximum phase, $F_{\theta MAX}$ (figure 4). The response of the TSM device in air and the impedance traces obtained with one side of the crystal immersed in a liquid are shown in figures 7a and 7b, respectively. Both the magnitude and the phase traces of the measurement in liquid have been broadened and the sharp edges and points have been rounded off, compared to those measured in air. These changes can be explained in terms of the equivalent circuit parameters, the bulk properties of the liquid and phenomena occurring at the solid-liquid interface.

Effect of the Electrostatic Capacitance, C_0

The TSM device behaves as a simple parallel-plate capacitor at off-resonance frequencies. The capacitance is related to the area of the electrodes, the thickness of the resonator and the dielectric constant of the quartz (equation 5). Since the dielectric constant is determined by the direction of the field, it must be corrected for the fringing of the field, an artefact arising from the finite size of the electrodes. Mounting stress will also affect the dielectric constant thus this also must be taken into account. Other factors that will affect C_0 include double-layer formation in polar or conductive solutions and any charge distribution on the surface of the electrodes.

32
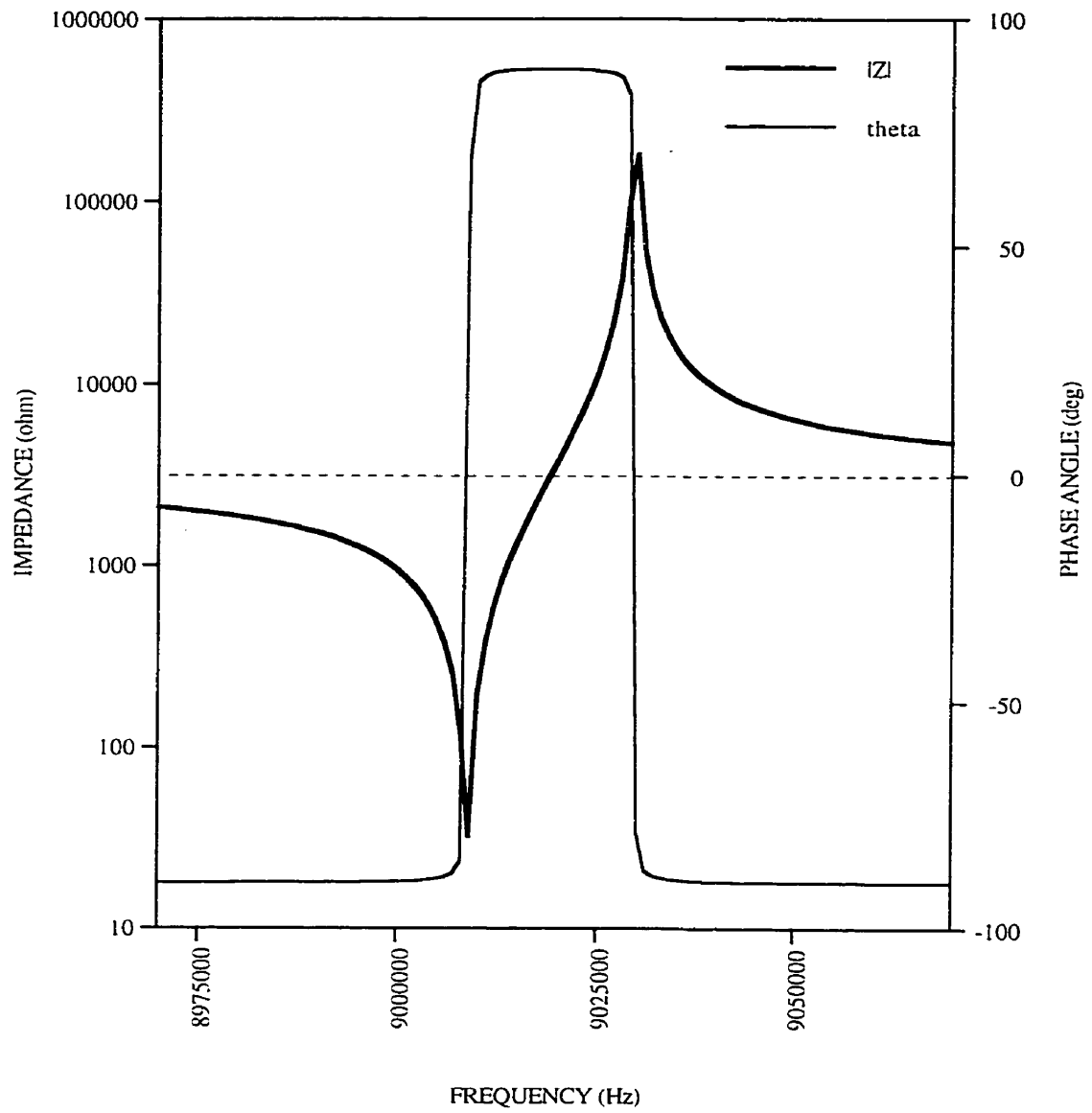
Figure 7a: Network analyser traces in air.
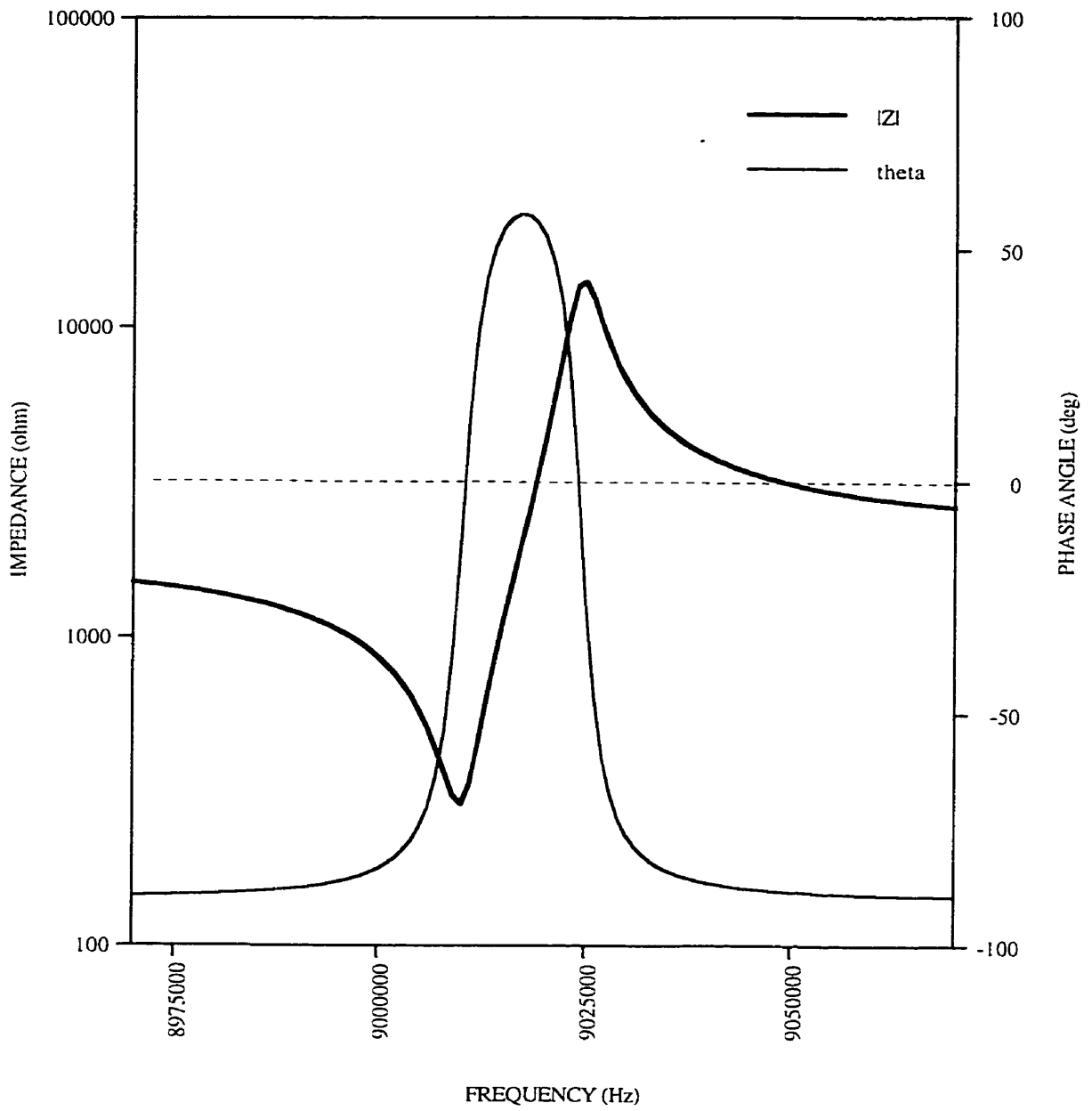
Figure 7b: Network analyser traces in liquid.

34

From equations 24, 26 and 28, it can be argued that F_R is independent of the electrostatic capacitance but that F_{ZMIN} and F_S should be dependent on C_o since both latter equations contain C_o. However, due to the high quality factor (Q) for most crystals (magnitude in the order of 10^4), a small change in C_o should have a minimum effect on both F_{ZMIN} and F_S. Similarly, from equations 25, 27 and 29, it can be seen that F_A, F_{ZMAX} and F_P are all dependent on the electrostatic capacitance. The presence of the extra C_o term in the expressions for F_{ZMAX} and F_P predict an inverse proportionality with C_o. These predicted effects of the electrostatic capacitance on the characteristic frequencies are illustrated in figures 8 and 9.[69] They depict network analysis simulations on a real measurement of a 9 MHz device in which C_o is varied while keeping all other parameters constant. Analysis of the impedance curve reveals that an increase in C_o causes a decrease in the frequency of maximum impedance while leaving the frequency of minimum impedance unchanged (figure 8). A similar effect is seen for the phase angle curve, in which F_P is reduced while F_S is unaffected (figure 9). Overall, this has the effect of reducing the resonant bandwidth which is indicative of an increase of energy storage in the resonator. Although the parallel resonant F_P is dependent on the circuit elements, its measurement can still be very useful in determining the properties of the media surrounding the resonator.

Effect of the Motional Resistance, R_m

The expression for R_m, depicted in equation X, relates to the energy dissipation for an infinitely large, unperturbed TSM device. Assuming all the other parameters stay constant for a particular resonator, R_m is directly proportional to the damping constant, σ_d. This parameter is assumed to contain all mechanisms of energy dissipation including internal dissipation, surface friction, coupling to the surrounding medium and mounting losses. It was shown by Muramatsu et al.[70] that for a quartz
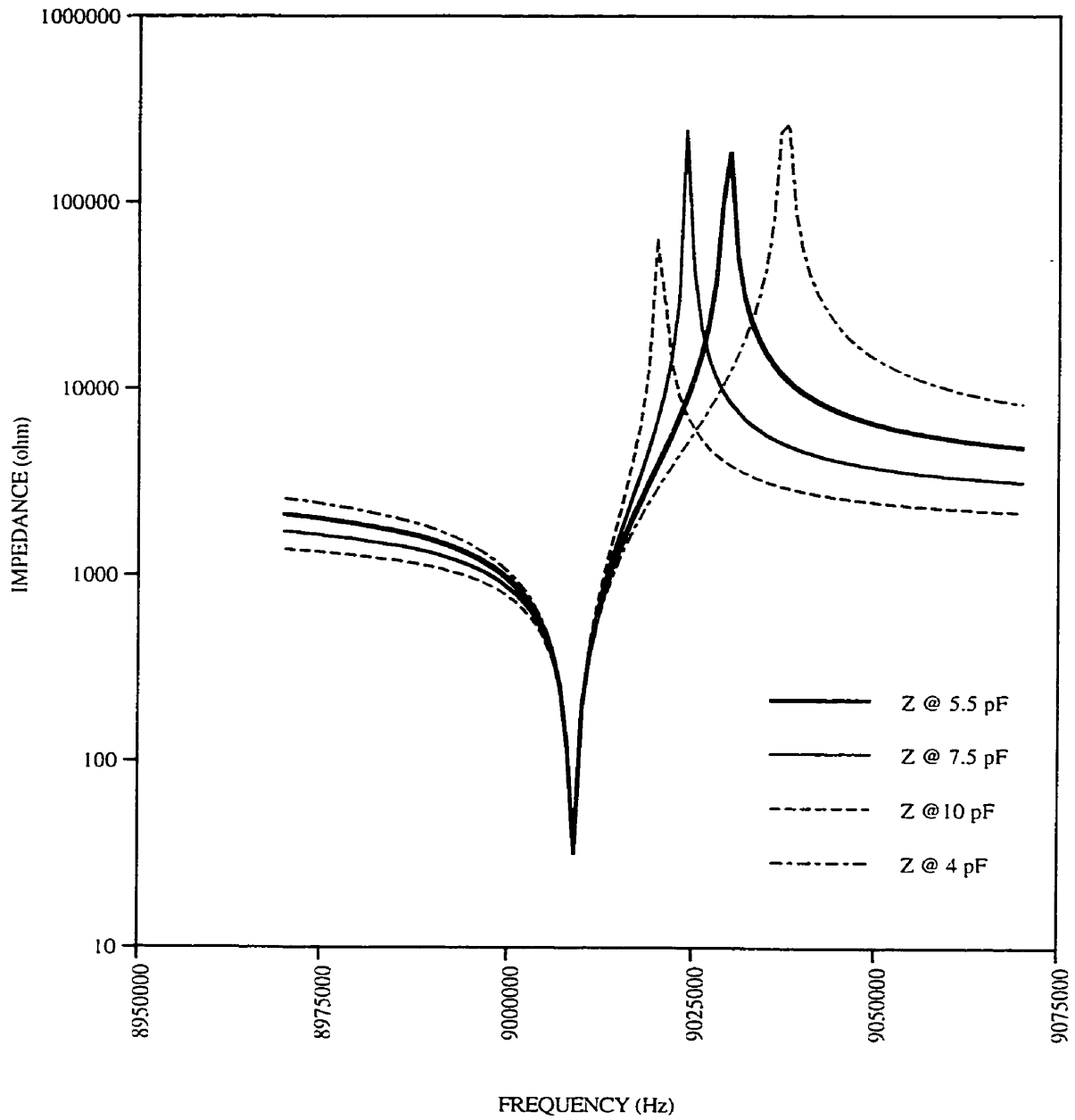
Figure 8: Effect of C_0 on the magnitude of the impedance.
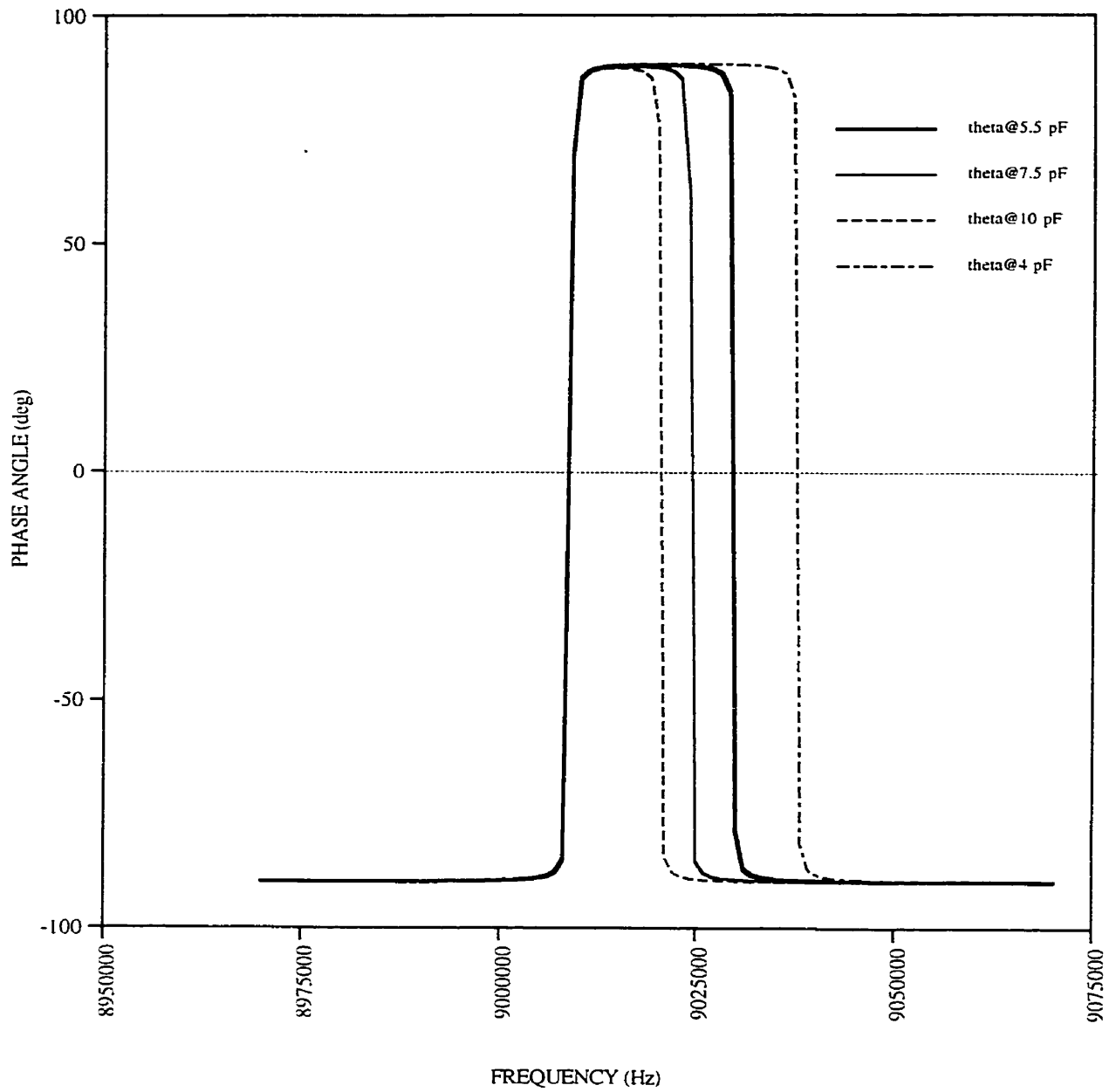
Figure 9: Effect of C_0 on the phase of the impedance.

37

crystal in contact with a liquid,

$$R_m = \frac{A}{K^2}(\omega\rho_L\eta_L)^{\frac{1}{2}}$$ (30)

where K is the electromechanical coupling factor, ρ_L is the liquid density and η_L is the liquid viscosity. However, this derivation is of limited use since the use of the coupling factor is arbitrary and its value cannot be determined experimentally. A later treatment of the TSM device through equivalent circuit analysis by Martin et al. [37] yielded a more useful expression for R_m.

$$R_m = \frac{\eta_Q}{\mu_Q C_1}\left(\frac{\omega}{\omega_s}\right)^2 + \frac{\omega_s L_1}{N\pi}\left(\frac{2\omega\rho_L\eta_L}{\mu_Q\rho_Q}\right)^{\frac{1}{2}}$$ (31)

where η_Q is the effective quartz viscosity, μ_Q is the quartz elastic constant, ρ_Q is the quartz density, C_1 is the unperturbed motional capacitance, L_1 the unperturbed motional inductance and N is the harmonic number with N = 1 for the fundamental mode. Unlike equation 30, this expression allows for the prediction of R_m from the properties of the quartz and the surrounding liquid. It can be seen from the above equation that Rm is dependent on both the density and viscosity of the coupled liquid thus the energy dissipation will also be proportional to the density-viscosity product.

Upon analyses of equations 24 and 25, it can be seen that both F_R and F_A are independent of R_m. The equations for F_S, F_P, F_{ZMIN} and F_{ZMAX} all contains a R_m term but once again. due to the high magnitude of the quality factor of quartz, the effects R_m has on all these characteristic frequencies should be minimal for small R_m increases. A similar manipulation of the R_m values for the experimentally derived

38
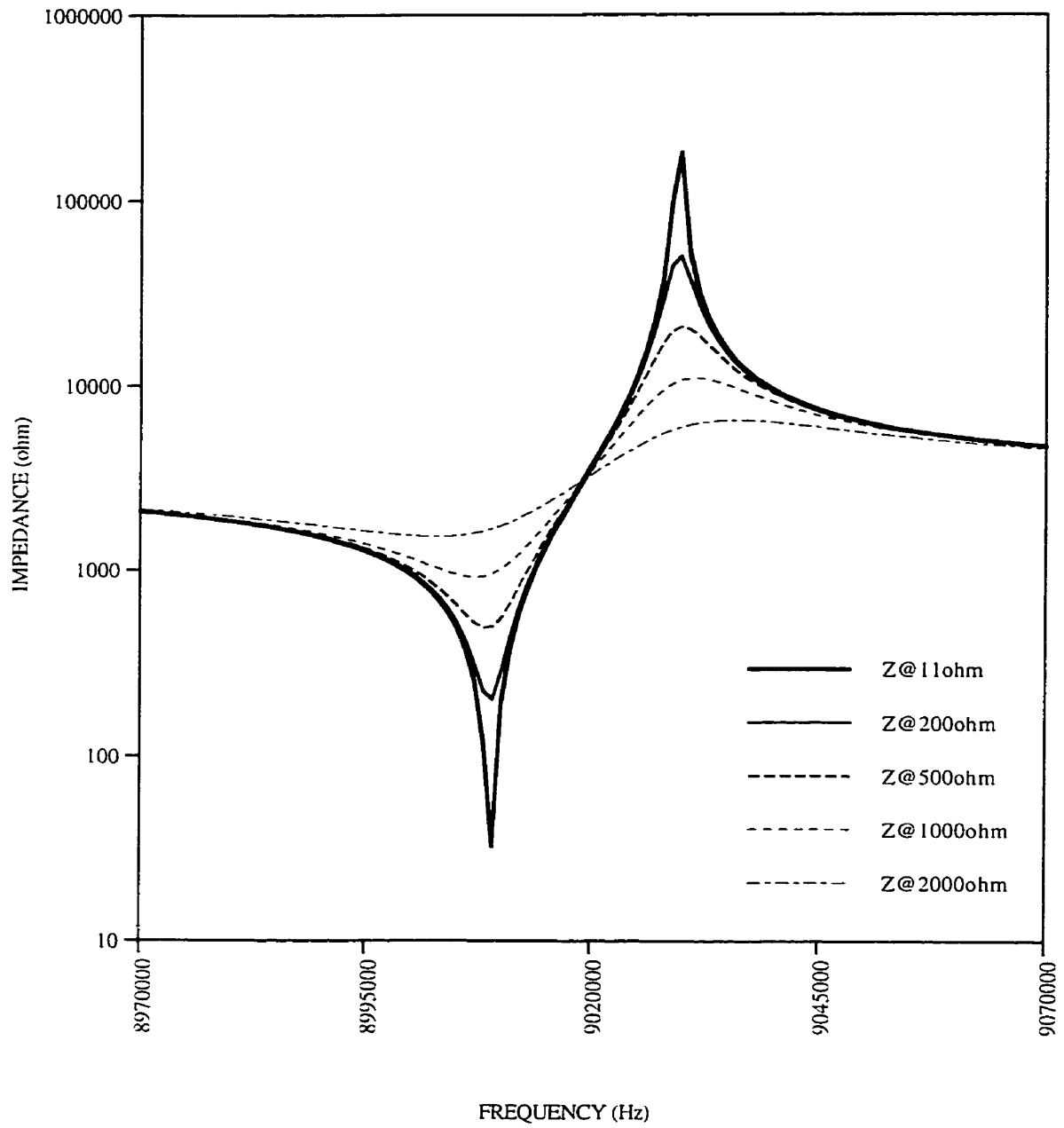
Figure 10: Effect of R_m on the magnitude of the impedance.

39
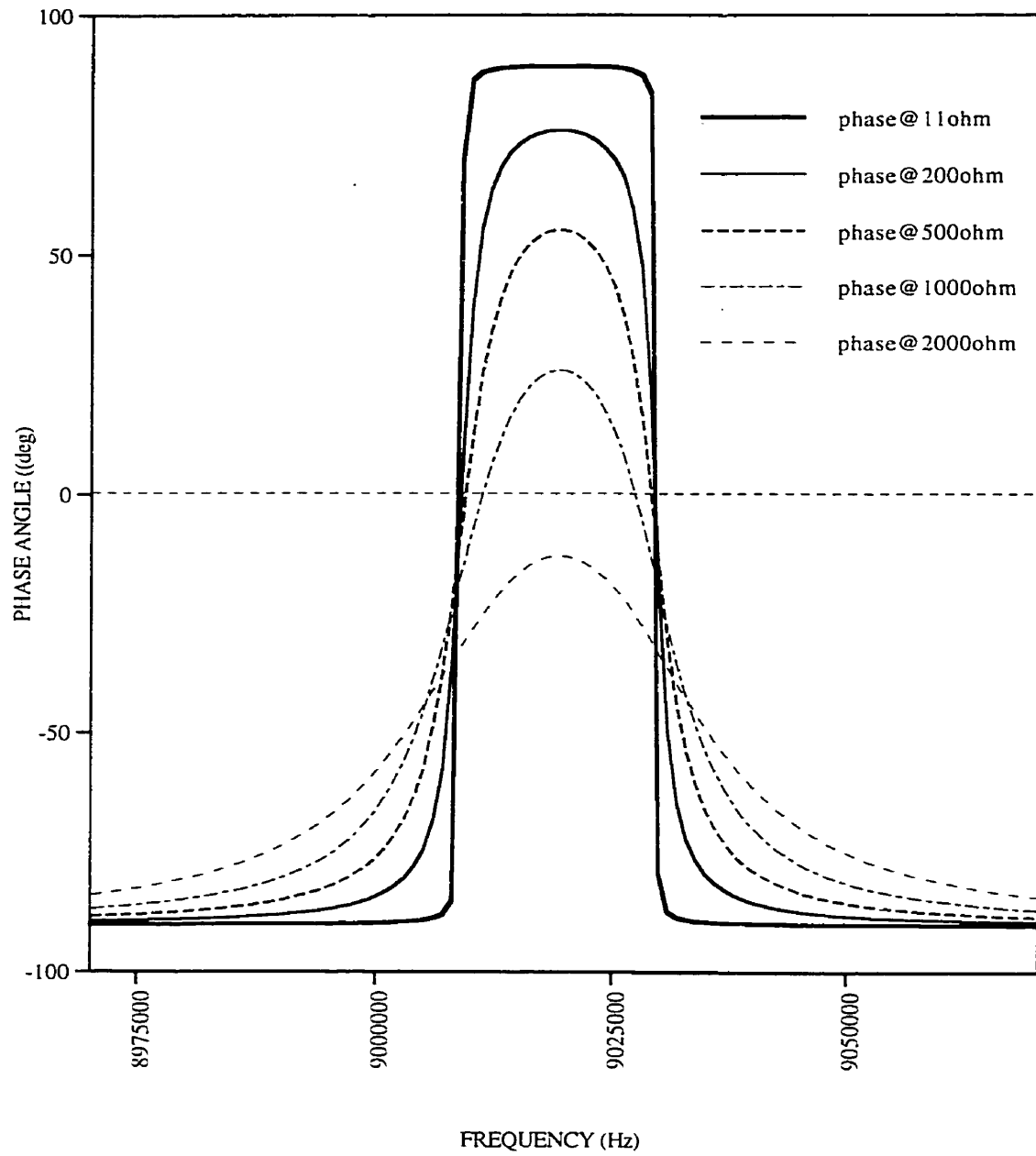
Figure 11: Effect of R_m on the phase of the impedance.

40

impedance curves is depicted in figures 10 and 11. An increase in R_m creates a damping, broadening effect on both the magnitude and the phase angle curves. This translates to an increase in the magnitude of the minimum impedance, Z_{MIN}, a decrease in the maximum impedance, Z_{MAX} (figure 10) and a decrease in the maximum phase angle, θ_{MAX} (figure 11). The characteristic frequencies, F_{ZMIN}, F_{ZMAX}, F_S and F_P are minimally affected for low R_m values, as predicted. However, the effects on these frequencies become more pronounced for larger R_m. It can be seen from the phase angle curve that the values for F_S and F_P converge with increasing R_m and eventually coincide at $R_m \approx 1500\ \Omega$, where the maximum phase angle, θ_{MAX}, becomes zero (figure 11). This critical value of motional resistance, R_m^*, above which both frequencies at zero phase cease to exist can be calculated by equating equations X and X and solving for R_m.

$$R_m^* = \frac{\sqrt{2}}{4\pi F_R C_0} \tag{32}$$

This cessation of resonance is one of the main limitations of the oscillator method that prevents its application to viscous liquid. The large number of parameters that are obtainable through network analysis allows for the characterisation of the quartz crystal even under this extreme condition as all other elements still exist, regardless of the value of R_m.

Effect of the Motional Inductance, L_m:

According to the Sauerbrey equation, changes in the resonant frequency can be related to mass changes at the surface of the TSM device. Such mass changes are also reflected in the value of the motional inductance, L_m since it is the inertial component

related to the displaced mass during oscillation. It represents the mass of the quartz

plate, any coating on the surface and also must be generalised to account for the

effects of the contacting fluid. A practical expression relating L_m to the properties of

the quartz and the coupled fluid was proposed by Lasky et al. [71]

$$L_m = L_1 + \frac{\omega_s L_1}{N\pi} \left(\frac{2\rho_L \eta_L}{\omega \mu_Q \rho_Q} \right)^{\frac{1}{2}} \tag{33}$$

where L_1 is the motional inductance of an unperturbed resonator with the second part

accounting for the liquid loading. Analyses of equations 24 to 29 reveal that the term

L_m is included in all the expressions for the characteristic frequencies in manners that

indicate a dependent relationship. This relationship is depicted in figures 12 and 13,

in which different values of L_m are imposed on the impedance measurements of an

unperturbed TSM device while keeping all other parameters constant. Increases in

the value of the inductance produce proportional decreases in all the characteristic

frequencies while keeping the values of Z_{MIN}, Z_{MAX} and θ_{MAX} fairly constant. The

fact that no changes in the resonant bandwidth is observed is further proof that

changes in L_m is purely mass dependent and has no effect on the energy dissipation

process of a resonator.

Effect of the Motional Capacitance, C_m:

The expression for the motional capacitance (equation 4), determined by

electromechanical analysis, relates to the elasticity of the quartz crystal. Apart from

the physical dimensions, an important variable in the equation is the elastic constant,

μ_Q, which corresponds to lattice restoring forces. Unlike the other circuit elements,

42
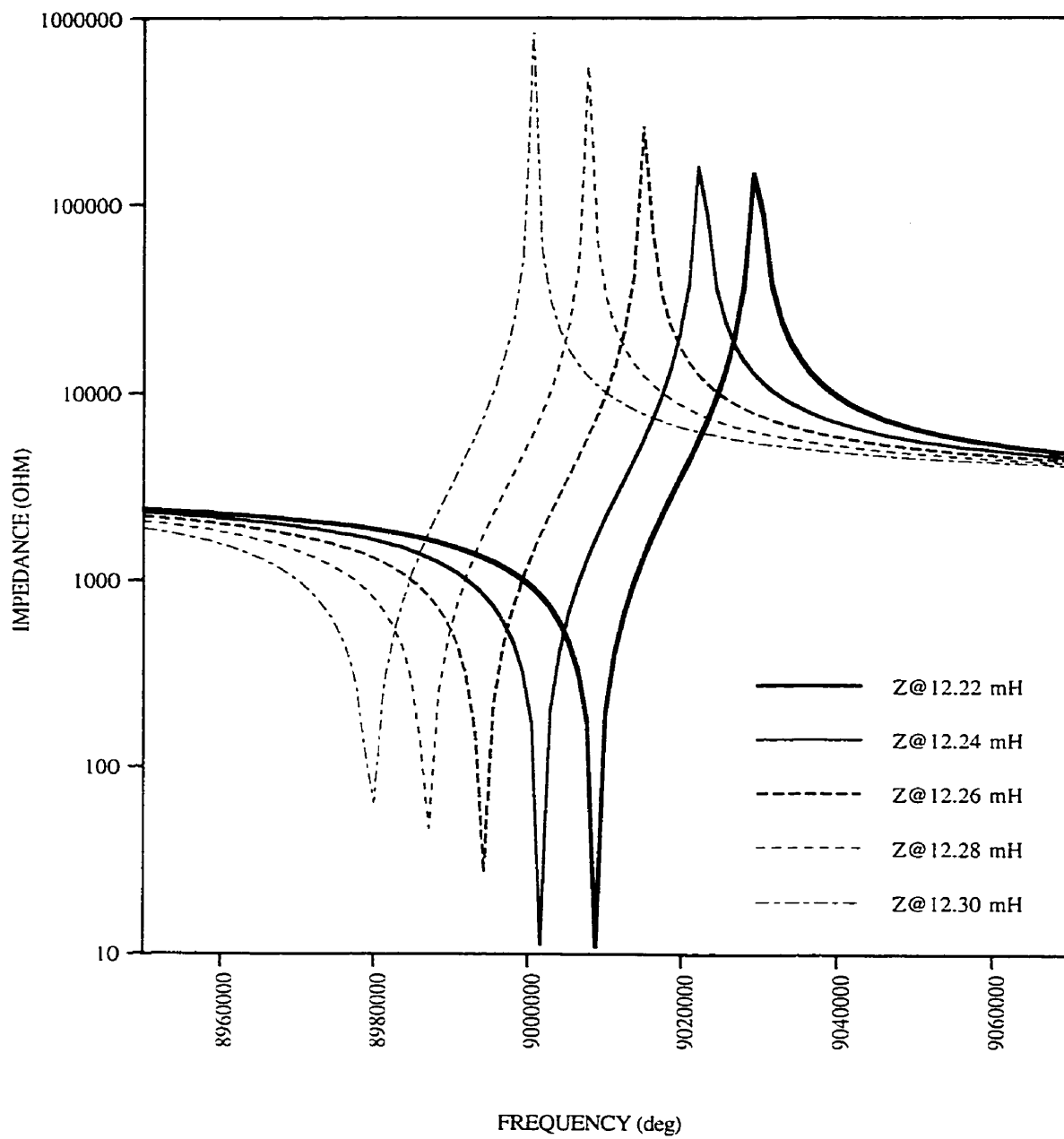
Figure 12: Effect of L_m on the magnitude of the impedance.
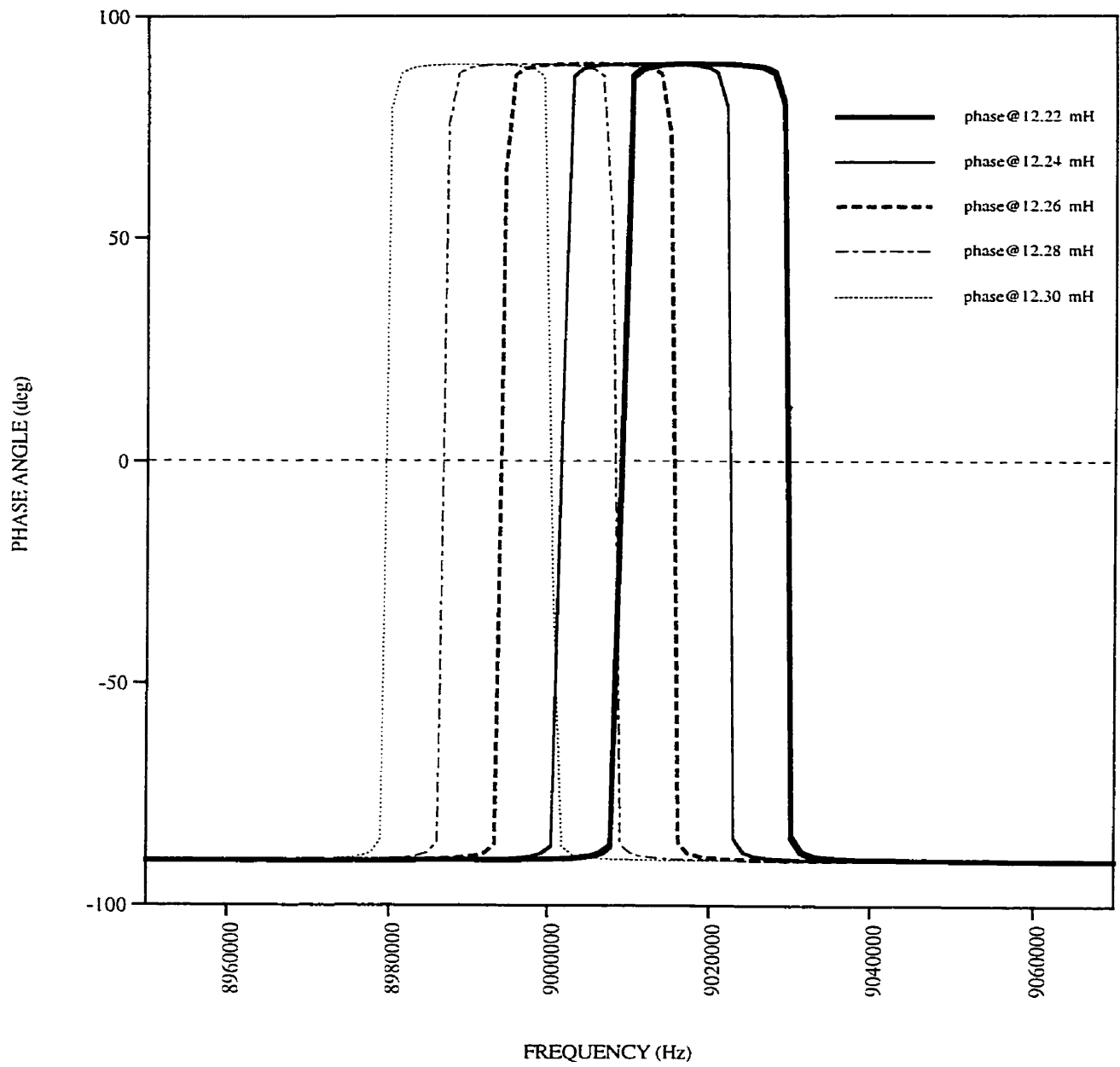
Figure 13: Effect of L_m on the phase of the impedance.

44

C_m has often been considered to be a constant under liquid loading since a Newtonian liquid does not possess any appreciable elasticity. In practice, the value of C_m does change in solution-based experiments but this has always been regarded as a possible artefact of the non-linear regression method used to fit the equivalent circuit elements to the experimental impedance curves. The effects that changes in Cm could exert on the impedance responses of the TSM device are shown in figure 14 and 15. An increase in Cm would shift all characteristic frequencies to lower values but would not have any effect on the magnitudes of the minimum impedance, the maximum impedance nor the maximum phase angle. These changes are very similar to those effected by a change in Lm. Thus, it is likely that changes that should be attributed to Lm may be wrongly, or disproportionally, attributed to Cm by the regression routine. Recently developed theories concerning the existence of non-Newtonian interfacial liquid structures that possess viscoelastic properties may confer intrinsic meaning to this variation in motional capacitance and this in turn, could give greater insight into interfacial phenomena.
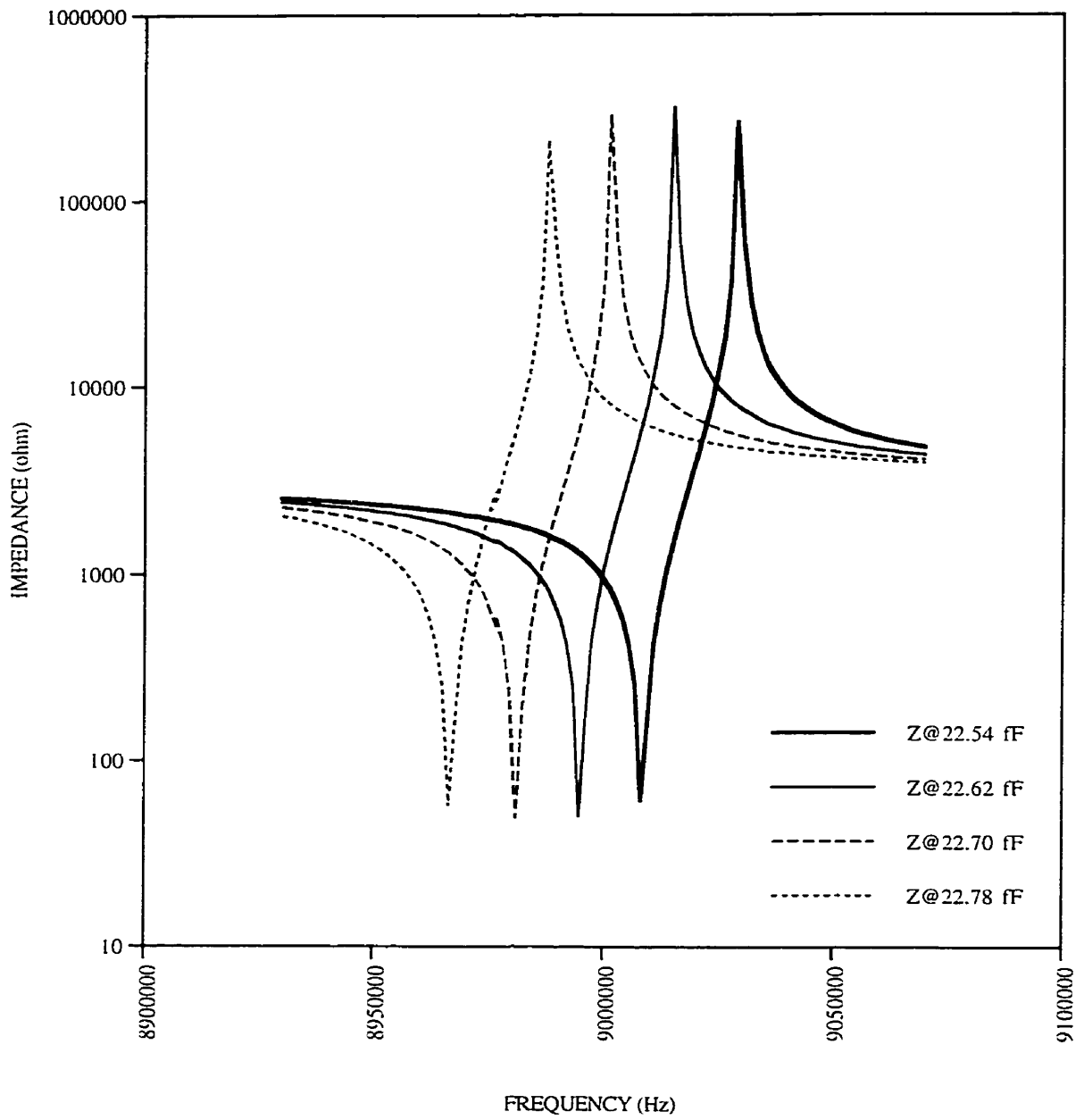
Figure 14: Effect of C_m on the magnitude of the impedance.
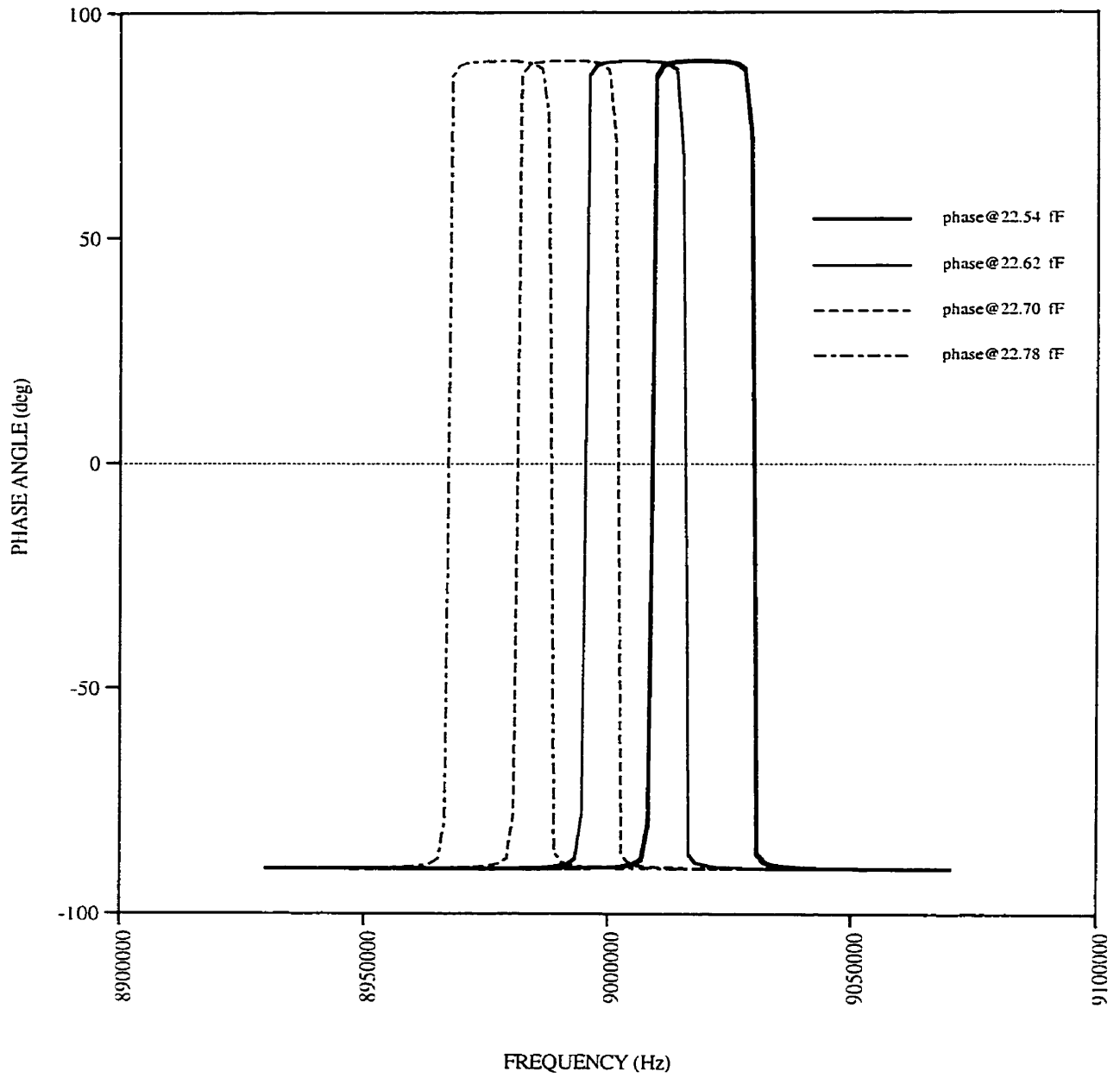
Figure 15: Effect of C_m on the phase of the impedance.

ARTIFICIAL NEURAL NETWORK THEORY

The human brain is the most complex computing device known to man. Artificial neural network (ANNs) arose from attempts to model its complex functions in a very fundamental manner. Neural network research got its start from the neuron model proposed by Mc Culloch and Pitts in 1943. [72] In 1958, Rosenblatt and co-workers developed perceptrons, which connect the above neurons in layers and used them to study pattern classification.[73] For a two-layer perceptron, convergence of learning was theoretically proven, drawing much interest to the field. A boom in neural network research followed until 1969 when Minsky and Papert demonstrated an important limitation of perceptrons *i.e.* that they are only applicable in cases where data, belonging to different classes, are linearly separable.[74] Interest and funding rapidly shrank and the field was basically dormant until 1985, when Hopfield devised a mutual connected model which can solve a very difficult combinatorial optimisation problem efficiently.[75] A year later, Rumelhart and co-workers presented a multilayered network trained by a back-propagation algorithm.[76] Convergence of learning was demonstrated and the network was proven to be applicable to non-linear problems. A renaissance occurred leading to new models being developed, new applications invented with several neuralnet-based systems being in actual use today.

OVERVIEW OF ANN's

The structure of the basic processing element in a neural network (figure 16), a neuron (also termed neural node or neurode), is designed to mimic that of a biological neuron. Both biological and artificial neurons receive multiple signals from primary sources or from nearby elements via dendrites and input channels, respectively. The magnitudes of these signals are modulated by the synaptic strengths

48
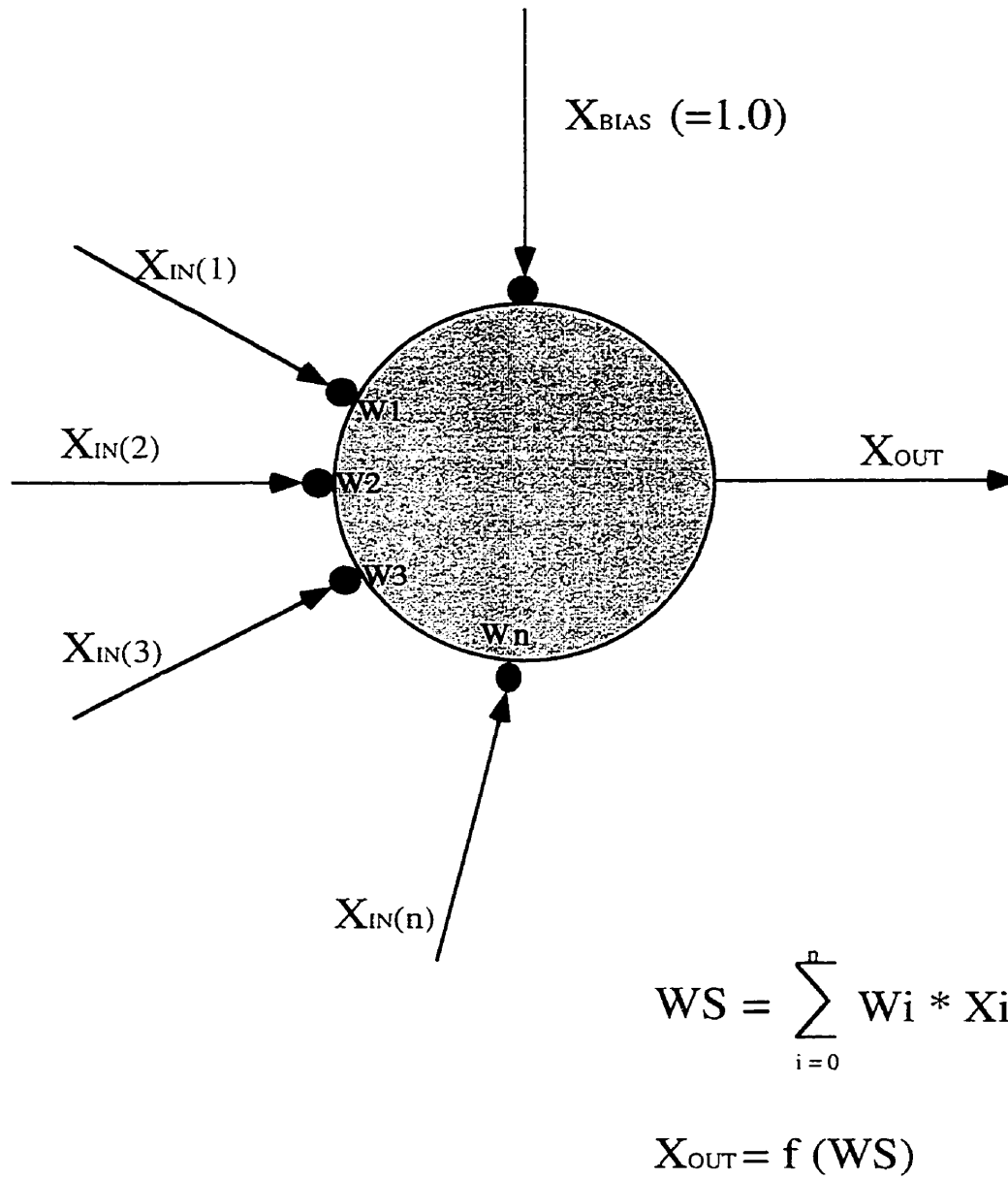
Figure 16: Structure and operation of a neural node.

biologically and by weighting factors artificially *i.e.* these 'weights' act to increase (excitatory) or decrease (inhibitory) the incoming signals. A large signal arriving through an inhibitory synapse may have the same or even less effect than a small signal entering via an excitatory synapse. In any case, all incoming signals are combined, usually by a simple summation, to produce a net input.

$$WS = x_1 w_1 + x_2 w_2 + ... + x_n w_n = \sum_n x_i w_i \qquad (34)^{77}$$

At this stage, all knowledge of the input pattern is lost. The details of each specific input are no longer available as only the net effect is recorded. This net input, in turn, is transformed by a transfer / activation function into an output signal. Biologically, if the combined input is greater than a threshold level, it activates the firing of the neuron, producing an output signal. The transfer function for an artificial neuron can be an analogous threshold function, in which an output is produced only if the internal activity level reaches a certain limit, or it can be a continuous function of the summed input. The transfer function acts as a mapping function **f** (·), producing the output **y** = **f** (net input). Some common activation functions are depicted in table 2. The output signal of a neuron is passed on via the axon and output channels to the dendrites and input channels of the next group of biological and artificial neurons, respectively.

By themselves, both types of neuron are not very interesting. The interesting effects result from the way these processing elements interact with each other. The brain consists of tens of billions of densely interconnected neurons. Neurons receive inputs from affector sensory cells or other neurons and send outputs to other neurons or effector organs such as muscles or glands. Only about 10% of the neurons are of the input (afferent) and output (efferent) varieties while the other 90% are
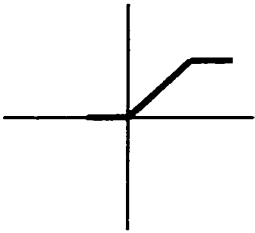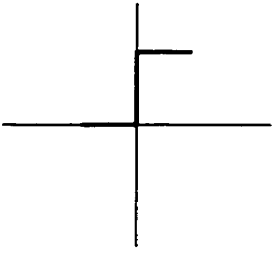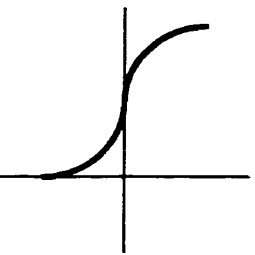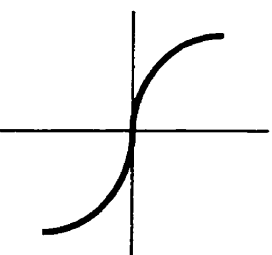
Function	Equation	Comments
 Linear	$F(x) = 0 \ (x \leq 0)$ $F(x) = mx$ $F(x) = 1 \ (x=1)$	-Linear, piecewise function. -Perceptron and ADALINE. -Not applicable to multilayer networks.
 Step	$F(x) = 0 \ (x \leq 0)$ $F(x) = 1 \ (x > 0)$	-Threshold function analogous to biological systems. -Piecewise and non-differentiable.
 Sigmoid	$F(x) = \dfrac{1}{1 + e^{-\beta x}}$	-Smooth version of a {0,1} step function. -Continuous and differentiable. -Applicable to nonlinear, multilayer network.
 Hyperbolic tangent	$F(x) = \dfrac{e^x - e^x}{e^x + e^{-x}}$	-Smooth version of a {-1,1} step function. -Continuous and differentiable. -Applicable to nonlinear, multilayer network.

Table 2. Common activation functions for neural networks

51

interconnected only to other neurons. The latter type are the ones responsible for the storage of information and for the transformations of the signals being propagated through the network. A single neuron may be connected to hundreds or even tens of thousands of other neurons via synapses. Neuronal activity is related to the generation of an internal electric potential which can be increased or inhibited by the input activity received from other neurons. If the cumulative inputs result in a potential above a certain threshold, the neuron fires a series of potential spikes along the axon to other neurons. These spikes initiate a signalling process which involves the release of a chemical neuro-transmitter at the terminating synapses. The strength of the signal depends on the amount of neuro-transmitter released by the axon and received by the dendrites. This synaptic efficiency, or strength, is modified when the brain 'learns'. It is believed also that some form of metabolic growth takes place in neurons as a result of increased cell activity. This growth process has been linked to learning and memory storage and may explain the difference in sizes (areas) of the synapses, which can vary by a factor of more than ten. Synapses with larger surface areas are believed to be excitatory while those with smaller areas are inhibitory.

An artificial neural network is also comprised of many interconnected elements but the sheer number is much smaller and the structure is greatly simplified. Like the brain, a suitable underlying structure is necessary for a neuralnet. It was determined empirically that the most appropriate arrangement is one in which processing elements are organised into interconnected layers, much as is found in the cortex and other parts of the brain. A typical network consists of several layers with full or random connectivity between successive layers (figure 17). There are two layers with connection to the outside world: an input layer where data is presented to the network and an output layer which holds the response of the network to a given input. The actual data transformation is done by the layers in between, termed the hidden layers.

52

RMS Error 0.0181

Output Layer

Hidden Layer

Bias

Input Layer
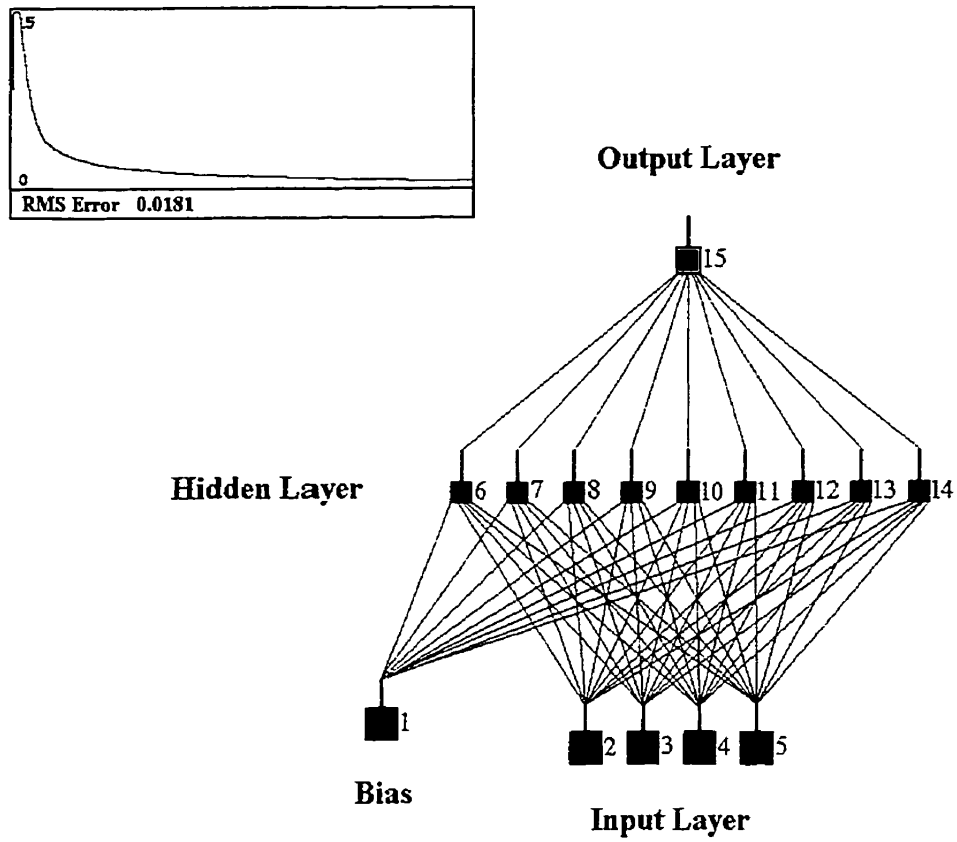
Figure 17. Architecture of a neural network extracted from NeuralWorks.

53

CHARACTERISTICS OF ANN'S

Before examining specific architecture, it is useful to review the general characteristics of ANN's that differentiate neural computing from the classical von Neumann serial architecture (digital computer) and parallel variations (expert system and artificial intelligence). Instead of a centralised memory storage device of a digital computer, information is stored, or more accurately, represented, in the pattern of variable interconnection weights among the neurons. Similarly, the computational process is a spreading, dynamic pattern of activity distributed across all the neurons in all the layers of a network. Unlike a computer, a neuralnet is taught or trained from examples rather than programmed. It can be regarded as a black box that transforms input vectors x, from a n-dimension space, to output vectors y, in a m-dimension space, using a function f, i.e. $f: x \rightarrow y$. The type of mapping a network can perform depends on the architecture, with more complex mappings generally requiring more elaborate architectures. Two categories of mapping exist: auto-associative, mapping to an original pattern from a noisy or partially given input pattern, and hetero-associative, mapping from an input pattern to a different output pattern.

The most important feature, which neural network shares with the human brain and differs from other computational methods, is the ability to generalise. Generalisation is the process of describing the whole from some of the parts, defining a class of objects from a knowledge of one or more instances or reasoning from the specific to the general case. This feature is essential to true learning, as opposed to merely remembering, as it allows the system to discern facts that apply to whole classes of objects rather than remembering many specific facts that apply only to individual members of each class. Generalisation is made possible by the fact that both neural network and the biological brain operate as a content-addressable

54

associative memory, a system which stores information by correlating it with other stored information. It is an extremely efficient mode of data storage and processing as without it, an unlimited number of specific events, facts and relationships has to be compartmentalised and still readily be available for recall. As a result of this, ANN's are able to compute or recall complete patterns from partial or noisy input patterns. This robustness in performance results from the fact that associative memories respond to the gross features of a stimulus and are generally not sidetracked by small differences or minor details. It is the overall input activity pattern that gives rise to the output pattern. An apt human analogy is the ability to recognise a familiar face in spite of a haircut or other cosmetic changes. Neuralnets also are able to classify objects not previously seen and to predict new outcomes from past behaviours. The ability to classify new objects is a form of interpolation among trained patterns while the predictive capacity is a form of extrapolation. Both types of mapping can be regarded as a form of generalisation.

Another advantage of neural computing arises as a direct consequence of its physical structure or architecture. Since the storage and computational processes are distributed among all weights and neurons, a neural network can be thought of as a massively parallel system. Like another massively parallel system, the brain, neural networks are highly fault tolerant. Whereas a complete system failure could arise from the loss of a single component in a serial computer, a neural network will keep working even after a significant fraction of its neurons and interconnections have become defective or inactive. This built-in redundancy is mirrored by the human brain. Thousands of brain cells die everyday with no appreciable decline in mental capacity and in extreme cases. portions of the brain can be damaged or removed without seriously affecting the performance of an individual. The massive parallelism of the biological brain is also responsible for the astounding speed at which it process

information. Consider the amount of computation required to process a single visual image having a resolution of 1000 x 1000 receptors, a modest number compared to the human retina. Over one million data points (much more for a colour image) must be examined and several million computations must be performed for the identification of such an image. Even at the nanosecond clock speed of modern digital computer, this process would require several seconds, perhaps more to ensure adequate recognition / identification. A biological visual system can complete such tasks in milliseconds, despite the fact that its signals propagate in millisecond rather than nanosecond time frames. This is due to the fact that these millions of computations are being performed simultaneously by the neurons of the biological neural system. Artificial neural networks are theoretically capable of the same performance once their architecture achieve the same complexity as that of biological neural networks.

OPERATION OF A NEURALNET

There are two main phases in the operation of a neural network: learning (or training) and recall. The learning process, while not identical to the biological model, is inspired by it. Similar to the brain, learning is the process of adapting or modifying the connective weights of the network in response to stimuli being presented at the input layer. Learning can be in the form of rote or inductive learning. Systems that learn by rote memorisation are not very interesting since they are just being programmed to become mere reference tables. On the other hand, inductive learning involves the formation of generalised concepts or rules from a number of specific instances. Generalisation is accomplished through the formation of classes of objects. This amounts to partitioning a universe of objects into two groups, those belonging to

the target class and those belonging to the complement class. In neural computing, the classes are defined by the mapping function, f, as boundaries in the weight space. The boundaries are formed through synaptic weight adjustments during the training process. Accurate generalisation occurs when the final class boundaries contain only the desired target examples and exclude all complement examples. The target examples include all patterns correctly belonging to the given concept classes, including those used in the training set and those not used.

Learning also could be supervised or unsupervised. In supervised learning, exemplary or desired outputs are presented to the network at the output layer to aid the weight updating process through comparison with the raw output generated by the network. The network will update the weights in a way to achieve the desired input / output mapping. Most supervised learning processes falls under one of three categories: Hebbian learning, gradient descent learning and stochastic learning. Hebbian learning is a rule, originally formulated to describe the human neural system, which states that a neural pathway is reinforced each time the activation on each side of the synapse is correlated. In artificial terms, this means that a connection weight on an input path to a neural node is incremented if both the input and the desired output is high. Gradient descent learning, an algorithm designed empirically for ANN's, is based on the reduction of the error between the actual output of a processing element and its desired output through the modification of the connection weights. A comparison is made between the desired output and the computed output to determine the error. The transfer function, f, is required to be differentiable since the iterative weight updates are based on the gradient of the error, which is defined in terms of the weights and the transfer function. The update rule is given as a solution to an equation such that

$$\Delta w_{ij} = \eta \frac{dE}{dw_{ij}} \qquad\qquad (35)^{78}$$

where η is a learning rate parameter and w_{ij} is the weight on the connection between neuron i and j. Lastly, in stochastic learning, the weights are adjusted in a probabilistic manner and the states of all units are determined by a probability distribution. During the learning phase, stochastic networks are operated in two modes: a clamped mode where the neurons are clamped to the values of the associate pairs of binary patterns and an unclamped mode in which there is no input. The network is allowed to reach an equilibrium at which the energy function of the system is at a minimum. At this stage, the weights of the network are adjusted on the basis of the difference between two state probabilities, clamped and unclamped.

In unsupervised learning, only inputs are provided to the network as stimuli and the learning process is totally dependent on the input pattern. Since there is no feedback on the desired or correct output, the system learns by discovering and adapting to features in the input patterns, such as statistical regularities or clusterings of patterns. Two learning rules have been utilised for this category, Hebbian learning and competitive learning. In competitive learning, the neurons compete among themselves and the one which yields the strongest response to a given input will have its weight adjusted favourably. There are many variations of this paradigm but most of the weight adjustment procedures are modified forms of Hebbian adjustment. In one variation, 'winner-takes-all', only the neuron with the highest activity within a layer is allowed to propagate its output. Also, this neuron will have its weight adjusted to resemble the input pattern while the weights of all other units remain

unchanged or get reduced. In other cases, the weights of neighbouring units are also strengthened to some degree. A third method of learning exists in which no specific output is provided for a given input but the network is graded on the quality of its response to a given input. Typically, the weights on units which give the right answer are reinforced while the weights on those giving the wrong answer are reduced. This is termed reinforcement learning and while interesting, it has not enjoyed the same utility and popularity as the other two paradigms.

Whatever kind of learning is used, the defining characteristic of any neuralnet is its learning rule. The learning rule specifies how the nodal weights adapt in response to exemplary inputs. Training may require showing a network many examples many thousands of times, or only once. The parameters governing a learning rule may change over time as the network progresses in its training. The long-term control of the learning parameters is referred to as a learning schedule. It is possible to overtrain a network such that a poor level generalisation is attained even though the error function has reached an acceptable level. An overtrained network had learned the training set too well and is thus unable to recognize target class patterns that were not part of the training set. Excessive training can occur for a limited training set that has been used repeatedly too many times in the training process. In attempting to learn accurately the individual pattern mappings, the target class boundaries are fitted too closely to the individual patterns. As a consequence some patterns, in the same class but not in the training set, are excluded from the boundaries, leading to poor generalisation. Overtraining can also occur when excessive neurons are used in a network for a given number of data points. The network 'memorises' the training set instead of forming a generalisation and in effect, becomes just a look-up table. The same phenomenon can be seen in polynomial function fitting in cases where the number of data points and the degree of polynomial are not chosen carefully. When

the degree of polynomial approaches the number of data points, the fit can be exact but rather than a smooth graph that follows the trend of most of the points, the curve follows the points themselves. This is termed overfitting and can be as invalid as underfitting (insufficient degree of polynomial).

Recall refers to how the network processes an input and generates a response at the output layer. For a trained network, recall is used to generate the required answers. In a supervised learning architecture, recall is an integral part of the learning process. It is used to create an actual output, which is compared to the exemplary output to generate an error signal. This error signal is incorporated into the learning rule to direct the weight modification process. For a particular training cycle, if a modification for a particular weight leads to a larger error, then either the magnitude or the direction of the change will be altered for the next cycle. At each step, the magnitude of the weight modulation is usually kept small to ensure that the network does not deviate too far from its partially evolved state. Although infinitesimal increments are the surest way to train a network, the consequence would be excessively large training times. This is why most learning schedule will vary the rate at which a network is modified in order to strike a balance between network stability and training duration.

NEURAL NETWORK ARCHITECTURES

The simplest type of neural network, in which all connections are from neurons of one layer to the next, is termed feedforward networks. These do not possess connections from one layer to the preceeding layer nor even among neurons of the same layer. Inputs are processed unidirectionally from input to hidden to output layer and any weight modification criteria calculated at the output are introduced to the

network in the next cycle. Feed forward networks can have as few as two layers, the input layer and the output layer. The earliest neuralnets are of this variety but they lack the ability to compute arbitrary complex functions such as those bounding nonlinearly separable regions. The solution to this problem was first proposed by Rumelhart et al. [76], a feedforward network having at least one hidden layer of neurons with each neuron capable of computing a nonlinear activation function. There is no theoretical limit on the number of hidden layers but typically there would be one or two. Investigations have been done which indicate that a maximum of three layers are required to solve complex pattern classification problems but no verification is possible. These multilayer feedforward networks are capable of performing any arbitrary mappings $f : R^n \rightarrow R^m$ if a sufficient number of hidden nodes are provided and if a set of weights that perform the desired mapping can be found.

Simple feedforward networks can be modified by connecting the output of one or more processing elements (PE) to the inputs of PE's in the same or even in the previous layer, yielding what is called recurrent networks. Within these networks, information will reverberate within layers and across layers until some convergence criterion is met. At this point, the state of the network is then passed to the output layer. The incorporation of such feedback connections will result in significant changes in the operation and learning processes of a neuralnet but the added complexity also imparts greater computational power over conventional networks. Unlike simple feedforward networks, recurrent networks exhibit dynamic behaviours and are able to perform mappings that are functions of time and/or space. The time and space sensitivities allow them to learn spatio-temporal pattern sequences that are characteristic of phoneme, spoken languages and sequential problem-solving methods. Since recurrent networks are basically reconfigured feedforward neuralnets, it is possible to derive a feedforward equivalence for any recurrent net. This is

61

accomplished through an unfolding-in-time process, in which each time step of a recurrent network is recreated by an additional layer in the feedforward analog. Although it will be able to perform the same function, the feedforward version would be significantly larger, thus the recurrent architecture merits and is generating substantial interest within the neural computing field.

The majority of neuralnets under investigation is still of the feedforward type. Arguably the most studied, the most utilised and thus the most successful paradigm is the backward error propagation or more commonly, the backpropagation network. It is also the paradigm of choice for this project thus a quick overview will be given with regards to its structure, characteristics and operation.

BACKPROPAGATION

The backpropagation network is a gradient descent method designed for a multilayer feedforward network with differentiable, nonlinear activation function. The first gradient descent algorithm, the delta rule, was designed for linear transfer functions. The update rule formulated for this algorithm (equation 35) will always find a set of weights which minimise the error for a set of linearly separable patterns. However, the delta rule cannot be applied to networks with hidden layers as it is unable to give credit or assign blame to the nodes in the hidden layers for errors that occur in the output layer. Since neither credit or blame can be assigned, the weights in the hidden layers cannot be updated in a logical manner. This is the 'credit assignment' problem, posed by Minsky and Papert in 1969 [74] and solved by Rumelhart et al. in 1985 [76]. Backpropagation overcomes this problem by assuming that all processing elements are somewhat to blame for the computed error. A learning cycle for a backpropagation neuralnet is divided into two distinct phases. In

the first phase, the training instances are propagated forward layer by layer until an output pattern is computed. In the second phase, the error at the output layer is calculated, through comparison with the desired output pattern. This then, is fed backward to the previous layer, the last hidden layer, to adjust the weights in that layer. Likewise, the error value calculated from the output of the last hidden layer is used to adjust the weights of the second-to-last hidden layer. This process is repeated until the input layer is reached and once all the weights in the network have been adjusted, the first phase of the next training cycle can continue anew.

A weight updating rule, analogous to the delta rule, can be derived for the nonlinear backpropagation network. The derivation is simplified if a network is assumed to have only one hidden layer but the update rule can be generalised for networks with an arbitrary number hidden layers. The following parameters are assumed for the derivation: a bounded, differentiable activation function f, n input neurons, h hidden neurons, m output neurons and an n-dimension training set of P patterns. To avoid confusion, different notations are used for the different layers as seen in the following relationships.

$$H_j = \sum_i v_{ij} x_i \qquad i = 1, 2, ..., n \qquad (36)$$

$$I_k = \sum_j w_{kj} y_j \qquad j = 1, 2, ..., h \qquad (37)$$

$$y_j = f(H_j) \qquad j = 1, 2, ..., h \qquad (38)$$

$$z_k = f(I_k) \qquad k = 1, 2, ..., m \qquad (39)$$

$$E_{mean} = \frac{1}{P} \sum_{p=1}^{P} E^p \qquad p = 1, 2, ..., P \qquad (40)$$

63

where H_j is the net input to hidden node j, I_k is the net input to output node k, v_{ij} is the weight connecting input node i and hidden node j, w_{jk} is the weight connecting hidden node j and output node k, x_i is the output from input node i, y_j is the output from hidden node j, z_k is the output from output node k, E_{mean} is the mean system error and E^p is the error of training pattern p. The full expansion of the computed value at the output, z_k, with respect to the other parameters yields

$$z_k = f(I_k) = f\left(\sum_j w_{kj} y_j\right) = f\left[\sum w_{kj} f(H_j)\right]$$

$$= f\left[\sum_j w_{kj} f\left(\sum_i v_{ji} x_i\right)\right] \qquad (41)$$

The aim of the learning rule is the minimisation of E_{mean} through the reduction of the error for each training pattern, E^p. A weight updating procedure that adjusts the weights in proportion to a reduction in the error, relative to changes in the weights, is required. This is accomplished by adjusting the weights in proportion to the negative error gradient *i.e.* at step $(t + 1)$, the weight adjustment should be proportional to the derivative of the error value E^p, calculated on iteration t. This can be written as

$$\Delta W(t+1) = -\eta \frac{dE^p}{dW(t)} \qquad (42)$$

where η is the learning coefficient and

$$\frac{dE^p}{dW} = \left(dE^p/dv_{11}, dE^p/dv_{12}, \ldots, dE^p/dw_{11}, \ldots, dE^p/dw_{hm} \right) \qquad (43)$$

The gradient of the mean system error is then given by

$$\frac{dE_{mean}}{dW} = \frac{1}{P} \sum_{p=1}^{P} \frac{dE^p}{dW} \qquad (44)$$

Since there are inherent differences, different update rules must be derived for the input-to-hidden and hidden-to-output connections. The weights connecting the output layer to the hidden layer will be considered first. Equation 42 can be modified to yield

$$w_{kj}(t+1) = w_{kj}(t) + \Delta w_{kj} = -\eta \frac{dE^p}{dw_{kj}} \qquad (45)$$

The error functions E_{mean} and E^p are usually defined as the mean square error in most systems since it penalises large deviations more than small ones. The squared term also provides a differentiable function of the difference between the computed and the desired output.

$$E^p = \sum_{k=1}^{m} \left(t_k^p - z_k^p \right)^2 \qquad (46)$$

where t_k^p is the desired output from output node k for the input pattern p and z_k^p is the actual output from output node k for the input pattern p. Using the chain rule and

65

substituting in equation 37 gives

$$\frac{dE^P}{dw_{kj}} = \frac{dE^P}{dI_k}\frac{dI_k}{dw_{kj}} = \frac{dE^P}{dI_k}\left(\frac{d}{w_{kj}}\sum_j y_j w_{kj}\right) \tag{47}$$

The first term can be evaluated using the chain rule and substituting in the derivatives of equations 46 and 39.

$$\frac{dE^P}{dI_k} = \frac{dE^P}{dz_k}\frac{dz_k}{dI_k} = -(t_k - z_k)f'(I_k) \tag{48}$$

The second term of equation 46 can be computed directly.

$$\frac{d}{w_{kj}}\sum_j y_j w_{kj} = y_j \tag{49}$$

and by defining the error term for each output node as

$$\delta_k = (t_k - z_k)f'(I_k) \tag{50}$$

the weight update rule for the output-to-hidden connections can be written in the same form as the delta rule.

$$\Delta w_{kj} = -\eta\frac{dE}{dw_{kj}} = \eta\delta_k y_j \tag{51}$$

66.

The derivation of the update rule for the weights between the hidden layer and the input layer is slightly more complicated since there is no desired values from which errors can be computed. The errors from the output nodes must be used to perform the adjustment and this can be done through the use of the chain rule to relate these errors to the weights in question. Equation 41 illustrates how deeply these weights are embedded in the error function. The update rule should be in a form analogous to equation 51 and by using the chain rule

$$\Delta v_{ji} = -\eta \frac{dE}{dv_{ji}} = -\eta \frac{dE}{dH_j} \frac{dH_j}{dv_{ji}} \tag{52}$$

The first derivative can be solved using the chain rule again and substituting in the derivative of equation 38.

$$\frac{dE}{dH_j} = \frac{dE}{dy_j} \frac{dy_j}{dH_j} = \frac{dE}{dy_j} f'(H_j) \tag{53}$$

.

Substituting equation 41 into equation m and differentiating yields

$$\frac{dE}{dy_j} = \sum_k \frac{d\left(t_k - f\left(\sum_j w_{kj} y_j\right)\right)^2}{dy_j}$$

$$= -\sum_k (t_k - z_k) f'(I_k) w_{kj} \tag{54}$$

The second part of equation 52 can be evaluated directly after substituting in equation 36.

$$\frac{dH_j}{dv_{ji}} = \sum_i \frac{d}{dv_{ji}}\left(v_{ji}x_i\right) = x_i \tag{55}$$

Substituting equations 53, 54 and 55 into equation 52 gives the second weight updating rule.

$$\Delta v_{ji} = -\eta x_i f'(H_j)\sum_k \left(t_k - z_k\right)f'(I_k)w_{kj}$$

$$= \eta \delta_j x_i \tag{56}$$

where

$$\delta_j = f'(H_j)\sum \delta_k w_{kj} \tag{57}$$

and as above

$$\delta_k = \left(t_k - z_k\right)f'(I_k) \tag{50}$$

In summary, the complete operation of a backpropagation network is as followed. Prior to training, all the weights in the neuralnet are initialised to small, random, real values. The exemplary input pattern vectors x^p ($p = 1, 2, 3,..., P$) are presented to the network and a corresponding output pattern z^p is computed. In detail, each of the input vector are multiplied by the appropriate weights v_{ij} and summed to produce the net input, H_j, to the nodes in the hidden layer. A transfer function f is used to transform this net input to a signal y_j, the output of the jth node in the hidden layer. All signals y_j are multiplied by the respective weight w_{jk} and summed to produce the net hidden-layer output, I_k. At the output nodes, I_k is transformed by f into the output-layer output z_k. This output is compared to the desired / target value t_k and an error $(t_k - z_k)$, with $k = 1, 2, 3,..., m$, is calculated. The error is used to calculate the delta term, δ_k to be used in the error backpropagation process, defined by equations 45 and 51. As can be seen from the gradient descent rule in equation 51, each of the weights w_{kj} connected to the kth output node is adjusted according to the size and direction of the negative gradient on the error surface.

The error propagation for the set of hidden-to-input weights, v_{ji}, is slightly more complicated. Since the hidden nodes do not possess target output values, nodal errors cannot be computed directly. The new delta term δ_j, must be computed from the output-layer errors according to equation 57. This is then backpropagated to adjust each of the weights v_{ji} connected to the jth hidden node, according to equation 56. Once all the weights in the system are manipulated, the training cycle is repeated. This process continues iteratively until a set of weights, W, is found, which minimises the global error function over all the training pattern. Alternatively, the training process may be stopped at a pre-set threshold error value or once a prescribed total

number of training cycles has been reached.

Convergence in Backpropagation

The error surface for a network with nonlinear activation function is a complex surface with many local and global minima.[79] This multiplicity of minima arises from the symmetry of the weights and nodes in the network. For any given error value E, there are many permutations of weights which will result in the same value of E. Since some E values must be in the minima region, this results in the multiplicity. In addition, the error surface is far more irregular than the smooth parabolic bowl-shape surface of the linear delta-rule algorithm. It contains flat plateaux, steep crevices, shallow troughs and other irregular geometries. These serve to impede the search for the global minimum as the network can become stagnant in a flat valley or trapped in a local minimum. This may result in very long training time and non-convergence, respectively, thus several training strategies have been devised to optimise the rate of convergence.

Weight Initialisation - The choice of the initial weight values can affect the rate of convergence or even induce a convergence failure. If the initial weights all have the same value, all hidden units will receive near identical error updates since the backpropagated error value is proportional to the weights. All weight adjustments will be identical and the system will be trapped at an equilibrium point that keeps the weights from changing. It has been shown by Hirose *et al.* [80] that very small initial weight values may lead to non convergence. The network became trapped in local minima and cannot escape since the small weights only yield similarly small updates.

70

On the other hand if large weights are initially chosen, the activation function, f, may become saturated. This is because the weight update is dependent on the derivative of f, which approaches zero as f approaches its maximum. Large initial weights will yield a large f values and a near zero f' values, leading to insignificant weight updates and slow downhill progress. A remedy is the addition of a constant , β, to the derivative term so that it will never fall below this level. A value of $\beta = 0.1$ has been shown to produce significant improvements in convergence rates [81].

Learning Rate Coefficient - The learning factor η, in the weight update rule, determines the magnitude of the weight adjustments made at each iteration. Since the adjustment is a linear function of the partial derivative (equation 42), an assumption is made that the error surface is locally linear with the locality being defined by the size of the learning coefficient. In areas of high curvature, this assumption will not hold for large coefficient. Large step sizes near steep, narrow features such as minima may result in an overshooting of the minimum and may set up an oscillation about the minimum. Thus it is advisable to keep this factor low in these areas to avoid divergent behaviour. However, if the value of η is too small, the descent will progress in very small steps, leading to very slow learning especially for flat or nearly linear areas. It is clear that to ensure efficient convergence, the learning coefficient should be varied as the training progresses, with the step size being reduced as a minimum is approached. This is usually done by adding a momentum term to the update rule, in which a portion of the previous weight change is incorporated into the present weight change.

$$\Delta w_{kj}(t+1) = \eta \delta_k y_j + \alpha \Delta w_{kj}(t) \tag{58}$$

where α is the momentum term. The added term acts as a low-pass filter since general trends are reinforced while oscillatory behaviours tend to cancel themselves out. This allows an optimally low learning coefficient but with faster convergence rates.

Batch learning - Two different approaches can be taken in the weight-updating procedure for a network: the weight matrix can be adjusted incrementally or in batches. Incremental learning consists of successively updating the weights after each training instance, with each instance presented in a random manner. In batch mode, the errors are accumulated over the whole training set, for one iteration (an epoch), before a single update is performed. The incremental approach inherently imposes a stochastic component on the weight update routine, which may help the neuralnet escape from a local minimum. However, as mentioned earlier, this procedure may result in the network taking successive steps in opposite directions, impeding convergence in the process. Through summation, the batch mode does provide a more accurate estimate of the gradient vector but at the costs of extra computing time and memory storage demand. The choice between the two approaches is dependent on the nature of the data set. The incremental method has been shown to perform well for on-line process control applications [6] in which the composition of the training set is constantly modified while the batch method is more suitable for non-dynamic data set.

Miscellaneous Options - The introduction of random noise into a network may help to avoid local minima traps. The random noise shifts the location of the error function which permits the minimisation process to escape from a local minimum. Judicious addition of noise may also improve the generalisation ability of the network through the prevention of overfitting. Conversely, excessive or inopportune noise introduction may cause the process to miss a global minimum if the location of the error function is shifted too much or at the wrong moment.

Another strategy that has been used to improve the rate of convergence is the use of different error functions. Although the mean-square-error function defined in equation m is the most popular, it has many irregularities that make convergence difficult. Since the summation of quadratic terms describes an elliptic rather than circular contour, a gradient descent will not point directly in the direction of a minimum. This is because the error surface is steeper in one dimension, resulting in a combined vector that is shifted more toward that dimension. Other functions that have been used with success are a logarithmic function which yielded fewer local minima [82], a function based on the information theoretic entropy which resulted in faster convergence [83] and a cubic error function. [77]

Implementation of a Backpropagation Network

The architectural determination of a neural network is the most important criterion in a neural computation since it determines the 'trainability' and the generalisation capability of the network. This refers specifically to the numbers of hidden layers and neurodes, since the numbers of neurons for the input and output layers are determined by the nature of the exemplary data. The optimal structure of a neural network is highly application-specific and is dependent on the complexity of the desired

73

mapping, the number of training instances available and the desired accuracy. In most instances, the architecture of the network is derived empirically *i.e.* through trial and error, but there are two rough guidelines that can be applied to most networks. Since most present neuralnet applications are simulated on serial computer softwares instead of true parallel hardware networks, computational cost considerations demand that the optimal architecture be also the minimal architecture. Fortuitously, the first guideline states that a single hidden layer will be sufficient for most applications. A second hidden layer may be required for particular complex or irregular mappings and only extremely unusual applications are likely to have more than two hidden layers.

The chosen number of processing elements for the hidden layers should be large enough to effect the desired mapping but not so large that it 'memorises' the training set and loses the ability to generalise over the complete range of mappings. The optimal number must be determined through trial and error but a general guideline for a starting point, proposed by Widrow *et al.*[84], is that

$$h = \frac{P}{10(m + n)} \qquad (59)$$

where h is the number of neurodes for the first hidden layer, P is the number of training instances and m and n the numbers of outputs and inputs, respectively. The initial number of neurons in the second hidden layer, where necessary, should be about half of that in the first layer.

Although the numbers of neurons in the input and output layers are predicated on the nature of the exemplary data, the process is by no means trivial. The number of

input nodes is especially important since these should correspond to the independent vector variables. Only relevant and independent variables should be chosen as inputs as this is crucial to the success of the network. A relevant variable is one which contains information which will help the network to classify objects into the target class and the complement class while an independent variable is one that are not correlated to other variable (non-collinear) and thus does not contain redundant information. The selection process demands a good knowledge of the task domain and even then, some preliminary experimentations are usually necessary. The number of output neurons corresponds to the dependent variable and is usually apparent from the application specifications. For a classification problem such as a writing recognition application, the number of output nodes would be twenty six if only letters are considered and thirty six if numbers are included.

Another important consideration in the implementation of a neural network is the size and nature of the exemplary data set. The data set should be divided into three mutual exclusive groups: a training set, a testing set and a validation set. Although the training set must be exclusive, the validation set can be culled from the test set if necessary. If there are enough instances, the training set and the test set should be of equal sizes otherwise, the former must take precedent. The training set must be large enough to allow for a sufficient number of hidden nodes to perform the desired mapping. In addition to sizes, the training test and the test set should also have similar RMS errors for a network to be valid. A network architecture which trains well *i.e.* a low training RMS, but does not test well is one which has poor generalisation capability. In this case, new architectures must be examined until this important criterion is satisfied. Once a network has been trained and tested to a satisfactory level, it may be advisable to prune the structure before deployment. Pruning refers to removal of neurodes which have near constant output over the training set or outputs

75

which are nearly identical to other neurodes. The neurodes with near constant outputs can be replaced by a bias term in the weights of the neurodes in the next layer. Similarly, one of the identical-outputs neurodes can be removed and its weight simply added to the weight of the remaining neurode. Pruning has been shown to reduce the processing time at little or no cost to the performance of a network.

PROJECT OUTLINE

Since neural calculations are inherently complex and data-intensive, this study was purposely focused on a relatively simple system; the responses of the TSM device to an electrolyte, potassium chloride (KCl). In the first part of the investigation, a series of KCl solutions of known concentrations were characterised through network analysis. While the multidimensionality of network analysis can produce twelve different parameters from impedance and equivalent circuit analyses (figure 4), not all were selected to be used in the initial phase. The reasons for this are two-fold: simplicity of computation since the ANN analysis was to be simulated and suitability of the parameters. Preliminary neural network analyses on a small data set were done to select the most independent, non-collinear variables that are most relevant to the desired output, the concentration. The four chosen quantities were the series resonant frequency (F_S) which represents mass loading and other interfacial effects, the parallel resonant frequency (F_P) which characterises the capacitive loading effect, the motional resistance (R_m) which is related to the energy dissipation of the oscillating sensor and the static capacitance (C_0) which reflects changes in the interfacial capacitance.

The selected parameters were used as inputs in the architectural determination of an optimal network using a scaled-up data set. Once the optimal architecture was found, the trained and tested network was used to predict concentrations of various unknown solutions, given the respective inputs. The accuracy and the generalisation capability of the neuralnet were evaluated. In the second phase of the study, an interferent, hypothesised to affect only one of the parameters, is added to the unknown samples in order to test the robustness of the trained network. The adulterant of choice was ethanethiol (CH_3CH_2SH), which is postulated to affect

77

mainly the series resonant frequency (F_S) through its absorption to the surface of the gold electrode through the well-known Au-S interaction. It has been shown that R_m and C_0 are not affected by a change in surface mass since the deposited layer can be regarded as an extension of the material comprising the electrode.[37] Thus, power dissipation, reflected by R_m, and the dielectric properties of the quartz plate, reflected by C_0, should remain constant. The parallel resonant frequency, F_P, can be affected by mass deposition at the surface but it is not as sensitive to mass changes as F_S.[45] The network, previously trained on the unadulterated data set, was used to predict the concentrations of the contaminated solutions. Its performance was judged for robustness *i.e.* the ability to recall the correct pattern from partial or noisy inputs. A scheme, designed to improve the predictive ability on the adulterated solutions, was devised. This involved the manipulation of the weights associated with the affected variable, F_S, in an effort to account for the effect of the adulterant. The magnitude of the weights connecting the F_S input nodes to the hidden layer were adjusted downward to lessen the deleterious effect of the interferent and the performance of the adjusted network was evaluated. An analogous calibration scheme may also be possible for an array of single-output sensors, but this methodology is expected to be far more complex with respect to theory and application.

A similar study, involving a different adulterant, was performed to validate this calibration scheme. The new adulterant, glycerol, was designed to affect mainly R_m by changing the viscosity and density of the bulk fluid (equations 30 and 31). The series resonant frequency, F_S, can also be affected by the density-viscosity product but to a lesser extent while the other two parameters should not be affected significantly. The predictive abilities of the unadjusted and the adjusted networks were determined in order to evaluate the robustness of the neuralnet and the efficacy of the calibration scheme, respectively. Both adulterants were used in the next phase

78

of the project, multiple interferents corrected by a multi-points calibration scheme. The robustness of the network was evaluated for the dual-adulterants system and a two-point adjustment scheme was implemented and the results evaluated.

A parallel project was effected in which a chemically-selective platform was developed in preparation for the next step, the application of neural analysis and the accompanying calibration scheme to a real-world chemical sensing system. A self-assembled, monolayer-based system utilising bifunctional thiols (thiol one end, carboxylic acid the other) was the logical surface-activation strategy for the gold-electroded BAW transducer. Thiol systems of a single chain length, mixed chain lengths and single chain length mixed with alkylthiol spacers were investigated to determine the optimal surface activation. Amino-terminated, single-strand DNA twenty five base-pairs long (25-mer) was immobilised on the thiolated surface via carbodiimide chemistry. The reactivity of this functionalised surface with complementary and non-complementary DNA strands of the same length was characterised by network analysis.

EXPERIMENTAL

MATERIALS

AT-Cut quartz piezoelectric crystals coated with gold electrodes were obtained from Classic Frequency Control, Inc., Oklahoma City, OK, USA. The resonant frequencies of the crystals were ~ 9 MHz and the electrode surfaces were polished to < 1μm. Potassium chloride (KCl) and glycerol were procured from J.T. Baker Inc., Phillipsburg, NJ, USA. The salt was dried at 120 degrees Celsius before usage. Doubly distilled, deionised water was used as the reference and as the solvent for all standard solutions. All thiols, bromide precursors, the anhydrous ethanol used as solvent in the self-assembly process and chemicals for the DNA immobilisation were purchased from Aldrich Chemical Company, Milwaukee, WIS, USA. Thiols, both purchased and synthesized, were distilled or recrystallised prior to usage while all other reagents were used as is. The amino-terminated precursor, phosphoramidites and various ultra-pure reagents used in the synthesis of the DNA single strands were purchased from Applied Biosystems, Foster City, CA, USA. DNA-synthesis grade acetonitrile was used in the DNA immobilisation procedure.

APPARATUS

DNA synthesis was performed using an Applied Biosystems 392 DNA/RNA synthesiser, Foster City, CA, USA. The TSM device was clamped in a cell with Viton O-rings on both sides (figure 18). One side of the crystal was kept under nitrogen while the other side was immersed in the analyte solution. Network analysis

Analyte out

N_2 gas out

Glass-filled Teflon Housing

O-Rings

Quartz crystal

To Network Analyser

Analyte in

N_2 gas in
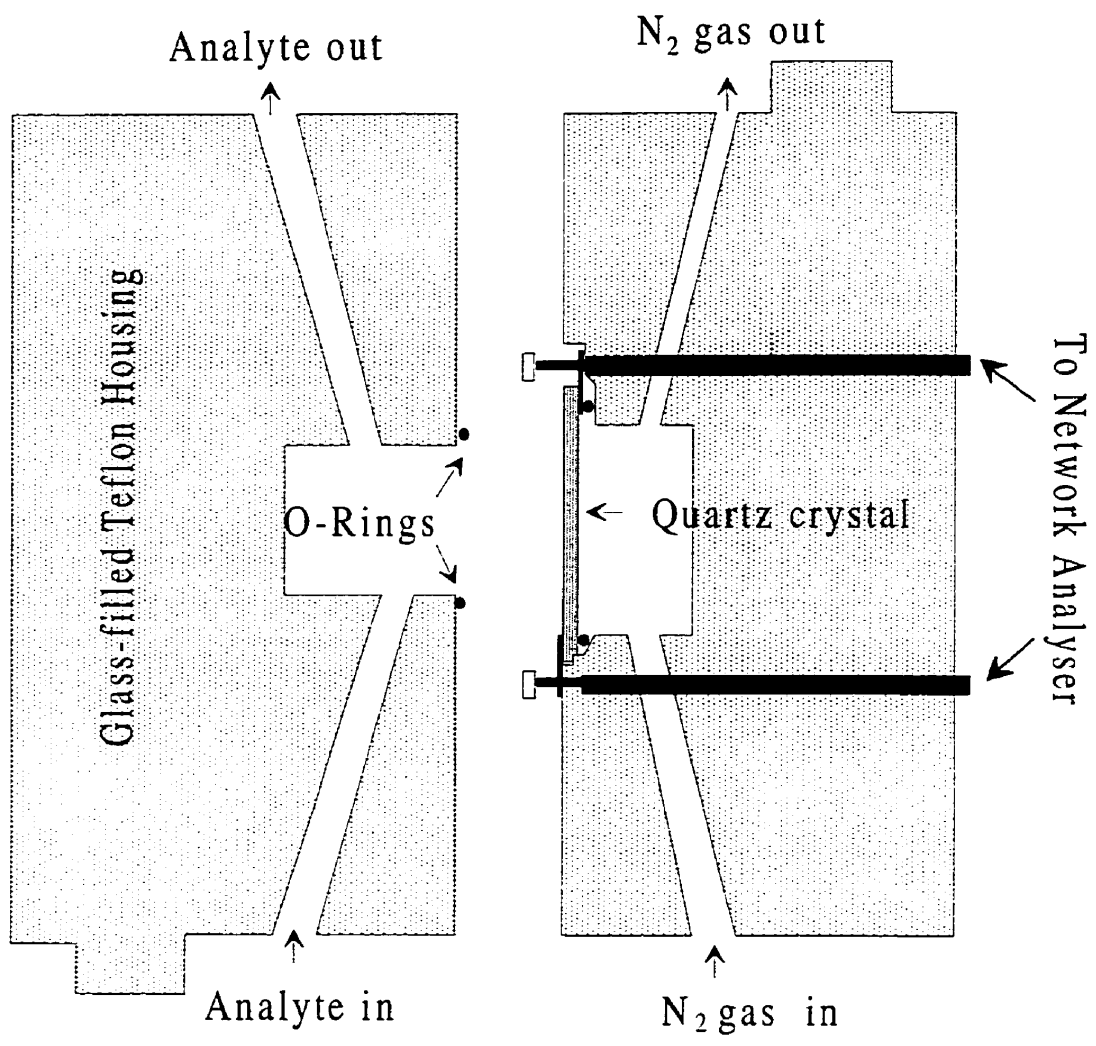
Figure 18: Flow cell .

was performed using a Hewlett Packard HP 4195A network/spectrum analyser, equipped with an HP 41951A impedance test kit and an HP 16092A spring clip fixture. A constant temperature circulator (model Hybaid 1267-62, Cole Palmer) was used to maintain the temperature for all analyses at 25 °C. Equivalent circuit analysis was performed by an HP proprietary algorithm built into the network analyser. All data were transferred to an IBM-compatible computer through an IEEE-488 interface from National Instruments, Austin, TX, USA and acquisition was controlled by an in-house C++ program running under a LabWindows platform (figure 19). A four-channel, peristaltic EVA pump, model 1000 was combined with an EVA injector, model 2000 (Eppendorf, Hamburg, GER) to introduce the analytical solutions to the cell.

X-ray photoelectron spectroscopy (XPS) spectral data were obtained with a Leybold MAX-200 spectrometer equipped with an unmonochromatic Mg Kα source, a concentric hemispherical analyser operating in constant transmission mode and a 18-channel detector (Leybold AG, Cologne, GER). Data were collected from a 1mm x 1 mm area with a 90° detector-to-sample alignment. AFM images were acquired using a NanoScope III system (Digital Instruments Inc., Santa Barbara, CA), utilising a J-type scanner (maximum scan size 120 x120 μm) and Nanoprobe silicon tips (Digital Instruments, Inc., Santa Barbara, CA) of 125 μm in length and 5-10 nm in radius were employed. All images were acquired under ambient conditions using the tapping mode of operation. In this mode, a sharp probe is brought into proximity with the surface of the specimen and is oscillated vertically near its mechanical resonance frequency (~ 300 kHz). Height and phase imaging were acquired simultaneously to maximise surface characterisation. The height mode displays a topographical relief of the surface while phase imaging depicts surface heterogeneities.

PUMP

FLOW CELL

ANALYTE

TO WASTE

NEURAL PROCESSING

HP4195A NETWORK ANALYZER
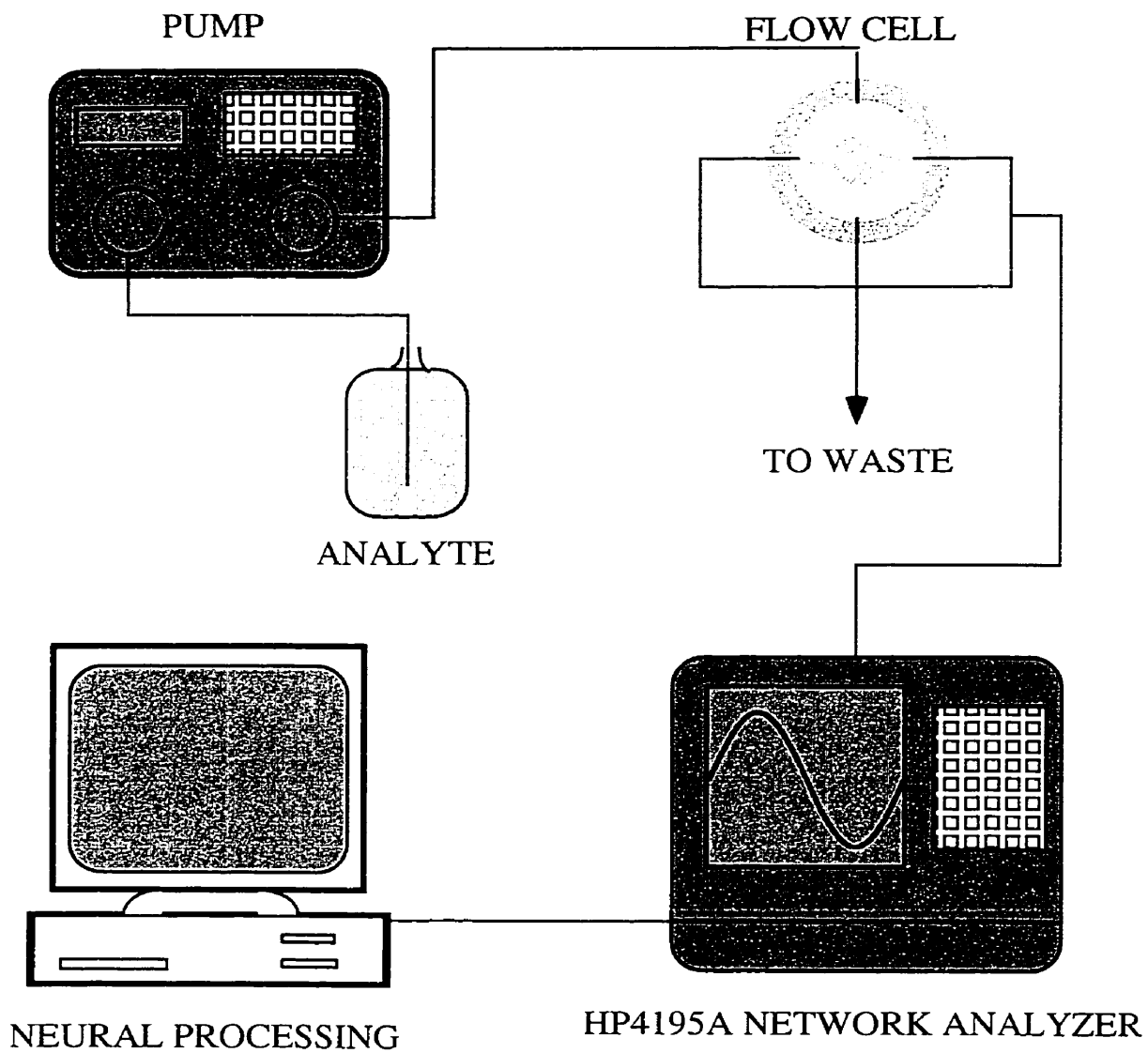
Figure 19: Instrumental setup.

Two different neural network packages were utilised in this project. The first, ANNAT, originates from collaborators in Åbo Akademi, Finland. This program uses the Levenberg-Marquardt algorithm for optimisation that makes use of an interpolation between the steepest descent and the Newton - Raphson methods. The Newton - Raphson method is a second-order optimisation routine, based on the approximation of the Hessian matrix of partial second derivatives of the error function. Second-order methods typically have convergence rates one order of magnitude higher than those for the gradient method (first partial derivatives). The Levenberg-Marquardt algorithm makes use of an adaptive parameter, λ, to effect a smooth interpolation between the gradient direction and the Newton-Raphson direction. The former is used far from the minimum, switching continuously to the latter as the minimum is approached. The software is robust, simple-to-use and faster than straight gradient descent backpropagation but it does possess some drawbacks, mainly that the number of weights (connections) should be kept below one hundred and that the architecture of a network must have the same number of nodes in each hidden layer. The second program is a commercial package, NeuralWorks Professional II/Plus v. 5.22, from NeuralWare, Pittsburgh, PA, USA. Several algorithms for RMS optimisation are offered but only back-propagation was utilised. This package does not possess the limitations of the in-house program but the typical back-propagation disadvantages apply here, that is, longer training and testing times and a tendency towards local minima.

PROCEDURES

Synthetic procedures for the thiols to be used in the modification of the gold electrodes were adapted from Whitesides and co-workers. [85] Thiolacetic acid was replaced by a better nucleophile, potassium thioacetate, to facilitate the reactions and increase the yields.

Synthesis of 1-undecanethiol - 4.5 mL of 1-bromoundecane (FW 235.22 g, 20 mmol) was combined with 2.15 mL (FW 76.12 g, 30 mmol) of thiolacetic acid ($CH_3COS^-H^+$) and 0.7 g of sodium metal (FW 22.99 g, 30 mmol) in 100 mL of degassed methanol and refluxed under an inert atmosphere for 2 hrs. 1.1 g of NaOMe (FW 54.02 g, 20 mmol) was added and the mixture was refluxed for a further 2 hrs. The resulting solution was cooled, quenched with degassed ammonium chloride solution and extracted with ether. The organic fraction was washed with water and saturated NaCl. After drying with $MgSO_4$, rotor vaporisation yielded 3.2 g of the 1-undecanthiol. Flash chromatography with hexanes produced 2.5 g of NMR pure product. ^1H NMR in $CDCl_3$ gives 2.5 (q, 2H), 1.5 (m, 2H), 1.3 (t, 1H), 1.0-1.5 (m, 16H) and 0.8 (t, 3H).

Synthesis of 6-mercapto-1-hexanoic acid - 9.27 g of 6-bromo-1-hexanoic acid (FW 195.06 g, 47.5 mmol) was combined with 10.86 g (FW 114.21 g, 95 mmol) of potassium thioacetate ($CH_3COS^-K^+$) in 200 mL of degassed methanol and refluxed under an inert atmosphere for 6 hrs. 50 mL of 1:1 water:conc. HCl was added to the cooled solution and the mixture was refluxed overnight. Extraction with diethyl ether

yielded a yellow liquid, confirmed by NMR to be the methyl ester of the thiol. This was refluxed with 50 mL of 1M NaOH in 100 mL of degassed methanol for 2 hrs. Acidification and extraction with ether gave a yellow liquid which solidified into a whitish slurry. Purification via flash chromatography with 4:1 ether:hexanes produced 6.5 g of the yellow liquid thiol and some white solid disulphide. ^{1}H NMR in CDCl$_3$ gives 2.5 (q, 2H), 2.3 (t, 2H), 1.6 (m, 4H), 1.4 (m, 2H) and 1.3 (t, 1H).

Synthesis of 8-mercapto-1-octanoic acid - 5 g of 8-bromo-1-octanoic acid (FW 223.12 g, 22.4 mmol) was combined with 5.2 g (FW 114.21 g, 45.6 mmol) of potassium thioacetate (CH$_3$COS$^-$K$^+$) in 200 mL of degassed methanol and refluxed under an inert atmosphere for 6 hrs. 50 mL of 1:1 water:conc. HCl was added to the cooled solution and the mixture was refluxed overnight. Extraction with diethyl ether produced a yellow liquid, the methyl ester of the thiol. This was refluxed with 50 mL of 1M NaOH in 100 mL of degassed methanol for 4 hrs. Acidification and extraction with ether gave a clear yellow liquid which solidified into a whitish slurry after vacuum pumping. Purification via flash chromatography with 4.5:1 ether:hexanes produced 6.5 g of the yellow liquid thiol and a minute amount of white solid disulphide. ^{1}H NMR in CDCl$_3$ gives 2.5 (q, 2H), 2.3 (t, 2H), 1.6 (m, 4H) and 1.4 (m, 6H).

Synthesis of 11-mercapto-1-undecanoic acid - 15 g of 11-bromo-1-undecanoic acid (FW 265.2 g, 56.6 mmol) was combined with 12.9 g (FW 114.21 g, 112.9 mmol) of potassium thioacetate (CH$_3$COS$^-$K$^+$) in 200 mL of degassed methanol and refluxed under an inert atmosphere for 8 hrs. 25 mL of conc. HCl was added to the cooled

solution and the mixture was refluxed overnight. Extraction with diethyl ether, followed by washes with sat. NaCl and water, yielded a yellow liquid, confirmed by NMR to be the methyl ester of the thiol. Hydrolysis with 50 mL of 1M NaOH in 100 mL of degassed methanol for 8 hrs followed by acidification and extraction with ether gave a dirty white solid. NMR showed that this was a 1:2 mixture of thiol:disulphide. Double recrystallisation from ethanol resulted in most of the disulphide crystallising out, leaving mainly thiol in the mother liquor. Further purification via flash chromatography with 2:1 ether:hexanes produced 4.8 g of the thiol in white powder form. ^{1}H NMR in CDCl$_3$ gives 2.5 (q, 2H), 2.3 (t, 2H), 1.6 (m, 4H) and 1.4 (m, 12H).

Synthesis of 16-mercapto-1-hexadecanoic acid - 5 g of 16-hydroxy-1-hexadecanoic acid (FW 272.43 g, 18.4 mmol) was combined with 200 mL of 1:1 48%HBr:glacial acetic acid and refluxed under an inert atmosphere for 2 days. Cooling of the mixture resulted in white crystals falling out of solution. The crystals were filtered, washed with cold water and recrystallised from hexanes, yielding 5.1 g of the 16-bromo-1-hexadecanoic acid. The bromide was converted to the thiol using the same procedure outlined above to give 4.2 g of white thiol powder. Recrystallisation from hexanes produced 2.8 g of NMR pure thiol in white powder form. ^{1}H NMR in CDCl$_3$ gives 2.52 (q, 2H), 2.35 (t, 2H), 1.6 (m, 4H) and 1.28 (m, 22H).

Self-Assembly of Thiols - Three different surfaces were prepared: a pure monolayer of a single chain length (C$_{16}$), a monolayer of mixed chain lengths (C$_6$, C$_8$, C$_{11}$ and C$_{16}$) and a single chain length (C$_{16}$) mixed with a shorter alkyl thiol spacer (C$_{11}$). All mixtures were comprised of equal proportions i.e. 1:1:1:1 for the mixed

bifunctional thiols and 1:1 for the C_{16} mixed with the C_{11} alkyl thiol spacer. The crystals were cleaned by soxhlet extraction with THF for up to 3 hrs prior to all immobilisation. An 80:20 mixture of ethanol and water was used as the solvent and thiol concentrations were kept in the 10 mM range, as per accepted protocols. [85] All crystals were immersed in the respective thiol solutions for 48 hrs after which they were thoroughly rinsed with fresh ethanol and cleaned by soxhlet extraction for up to 3 hrs to remove physisorbed impurities. Prepared crystals were kept in a dry dessicator until required.

Reactivity Assessment - Surfaces modified with the bifunctional thiols were reacted with a suitable probe molecule to determine their respective reactivities. Probe molecules containing multiple fluorine atoms were chosen to facilitate XPS characterisation due to the high sensitivity of XPS to fluorine. Thiolated surfaces with distal carboxylic groups (carboxylated surfaces) were reacted with pentafluorophenyl hydrazine (PFPH) in distilled dichloromethane, activated by 1,3 dicyclohexyl carbodiimide (DCC) and catalysed by 4-dimethylamino pyridine (DMAP). A control experiment was performed using the same procedure on an unmodified crystal. All modified surfaces were rinsed thoroughly with dichloromethane, cleaned by soxhlet extraction for 3 hrs and stored in a dessicator until required.

DNA Synthesis - Three different DNA sequences twenty five base-pairs in length (25-mer) were synthesised: an amino-terminated sequence intended for immobilisation (F1), the complementary sequence for hybridisation (F2) and a non-complementary sequence for control purpose (F0). The compositions of all three

sequences are depicted below.

F1:   5'-amino-TATAAAAAGAGAGAGAGATCGAGTC-3'

F2:   5'-   GACTCGATCTCTCTCTCTCTTTTTATA-3'

F0:   5'-   CCCTGAGCCACTATCAATCATGAG   -3'

The synthesised DNA sequences were purified with a Poly-Pak ion-exchange cartridge, containing a pH stable polymeric resin, designed for the purification of oligonucleotides synthesised by the DMT-on procedure. The pure DNA solutions were vacuum freeze-dried and stored at -5 °C until required.

DNA Immobilisation - The carboxylated quartz crystal was immersed in a solution of F1 DNA (1 mM), DCC (10 mM) and DMAP (< 1 mM), with acetonitrile as the solvent. The immobilisation was allowed to proceed for 24 hrs, follow which the crystal was rinsed well with acetonitrile, cleaned by soxhlet extraction for 3 hrs and stored in a dessicator until required.

Network Analysis - The network analyser was set to scan a frequency span of 40 kHz, centred about the resonant frequency of the crystal. A complete scan cycle, including equivalent circuit calculation, requires 10 seconds. A 50 seconds delay was programmed to yield a data acquisition frequency of one scan cycle per minute. All measurements were performed at a controlled 25 °C and the default solution flow rate was set at 0.07 mL/min. This ensured that the crystal always would be exposed to

89

fresh analytical solution.

For all KCl related measurements, doubly-distilled, deionised water was employed as the reference and as the solvent. The cell was filled with the reference, using a flow rate of 1 mL/min, following which the default flow rate was employed. The system was allowed to equilibrate for a short period of time, during which the series resonant frequency (F_S) was monitored. Data acquisition for the water reference would begin once F_S stabilises and this would be continued until 20+ data points had been collected. The flow stream then would be switched to the analyte and maintained there until the signal stabilises and 20+ stable data points are acquired. The flow stream would switch back to the water and a new reference data set would be collected to check for drift. This cycle would be repeated for any subsequent analytical solution. Analyses were performed for KCl concentrations between 0.001M and 1M. Adulterated solutions were KCl solutions, spanning the same concentration range, doped with sub-millimolar concentrations of ethanethiol (1 mM of thiol in the 1M KCl stock) and glycerol (1% v/v of glycerol in 1M KCl stock), respectively. Both the data for the water reference and the data for the analyte were averaged over the stable regime to arrive at mean values prior to any subsequent processing.

A Tris-buffer (pH 7.4) was used as the reference and as the solvent for all DNA-related measurements. Twenty-plus stable data points for the buffer would be collected at the default flow rate after which the flow stream would be switched to the complementary or the non-complementary DNA solution (0.25 mM). After enough stable data points have been acquired, the flow stream then would switched back to the buffer and a new set of data would be collected for the buffer. This cycle would be repeated for any subsequent DNA solution. Once again, the reference data set and

the analytical data set were averaged over the 20+ points stable regime, prior to subsequent processing.

Neural Computation - The initial phase of neural computation, determination of the optimal input parameters, was performed on ANNAT. Six parameters, determined from experimental data to be most relevant / sensitive to the output (KCl concentration), were chosen to form the input pool from which the four most independent will be chosen. An optimal architecture was determined for a system with 6 inputs and 1 output. The training error for this system was deemed to be the benchmark error, against which subsequent architectures were compared. Next, one input parameter was removed and the training RMS was determined. This exercise was repeated, each time removing one input parameter. The two parameters whose removal caused the least and the second least adverse changes in the training error, as compared to the benchmark error, were removed from the input pool. Since they contribute the least to the goodness-of-fit of the neuralnet, these two parameters can be considered to be less sensitive to the output and/or less independent than the others i.e. the information they contain is already present in one or more other parameter.

In the next phase, an optimal architecture was determined for the 4-inputs, 1-output system. For this stage, the test error was used in conjunction with the training error as criteria for the best design. The training and test sets contain 100 and 50 instances, respectively. Assessment of the predictive capability of the optimised network was performed using a validation set of 35 instances, culled from the test set.

The third phase involved the application of the trained network to the data set generated from KCl solutions doped with the interferent, ethanethiol. The network, trained on the unadulterated data set, was used to predict the concentrations of the

91

adulterated solutions in order to assess the robustness of the neuralnet. The magnitudes of the weights associated with the affected parameter, F_s, were adjusted downward incrementally until the testing error was minimised. The adjusted network was used to generate a new set of predictions for the adulterated solutions in order to evaluate the effectiveness of the weight adjustment procedure.

The fourth phase was analogous to the third phase, except that the adulterant was now glycerol and the affected parameter was the motional resistance, R_m. The network was applied to the data obtained from the doped solutions to test the robustness of the network with respect to glycerol contamination. The appropriate weight adjustments were made, new predictions were computed and the efficacy of the process was determined. A two-point calibration routine was also employed in an effort to reduce the predictive error. The adjusted network was used to predict the KCl concentrations of the doped solutions and the effectiveness of the two-points calibration procedure was assessed using the new predictions.

The fifth phase of the computation concerned the application of the network to solutions doped with both ethanethiol and glycerol. The robustness of the neuralnet with respect to the adulterants was evaluated, after which a two-points weight adjustment procedure was effected. Once again, the adjusted network was used to predict the KCl concentrations of the doped solutions and the effectiveness of the two-points calibration procedure was assessed using the new predictions.

RESULTS AND DISCUSSIONS

NEURAL COMPUTATION

Network Analyses of KCl System

Most applications of TSM devices in the liquid phase, especially in the biosensor and electrochemical fields, involve the use of conductive solutions. This and the fact that it is a relatively simple, well-studied system make the electrolyte, potassium chloride, the logical testbed for neural analysis. The responses of the TSM device to a series of KCl solutions of concentrations ranging from 0.001 M to 1 M were characterised through network analysis. Figures 20 to 23 depict the typical response patterns of various network analysis parameters, relative to water. These parameters exhibit substantial, reversible responses to the electrolyte, rendering them strong candidates as inputs in the neural computation for the KCl system. These responses can be explained in terms of the physical properties of the electrolytic solutions.

It has been demonstrated that for dilute conductive solutions, possessing density and viscosity close to those of the solvent, the response of the sensor cannot be explained adequately by just the mass and viscous effects.[40] In addition to mass loading and viscous coupling, the mechanisms of interaction between a TSM sensor and a solution may also include acoustoelectric coupling. Thus, the response of the sensor may be influenced by the electrical properties of the contacting solution. The propagation of an acoustic wave in the quartz crystal generates surface charges through the induced electrical polarisation of the crystal. This, in turn, creates an external electric field which extends into the surrounding fluid. The interaction
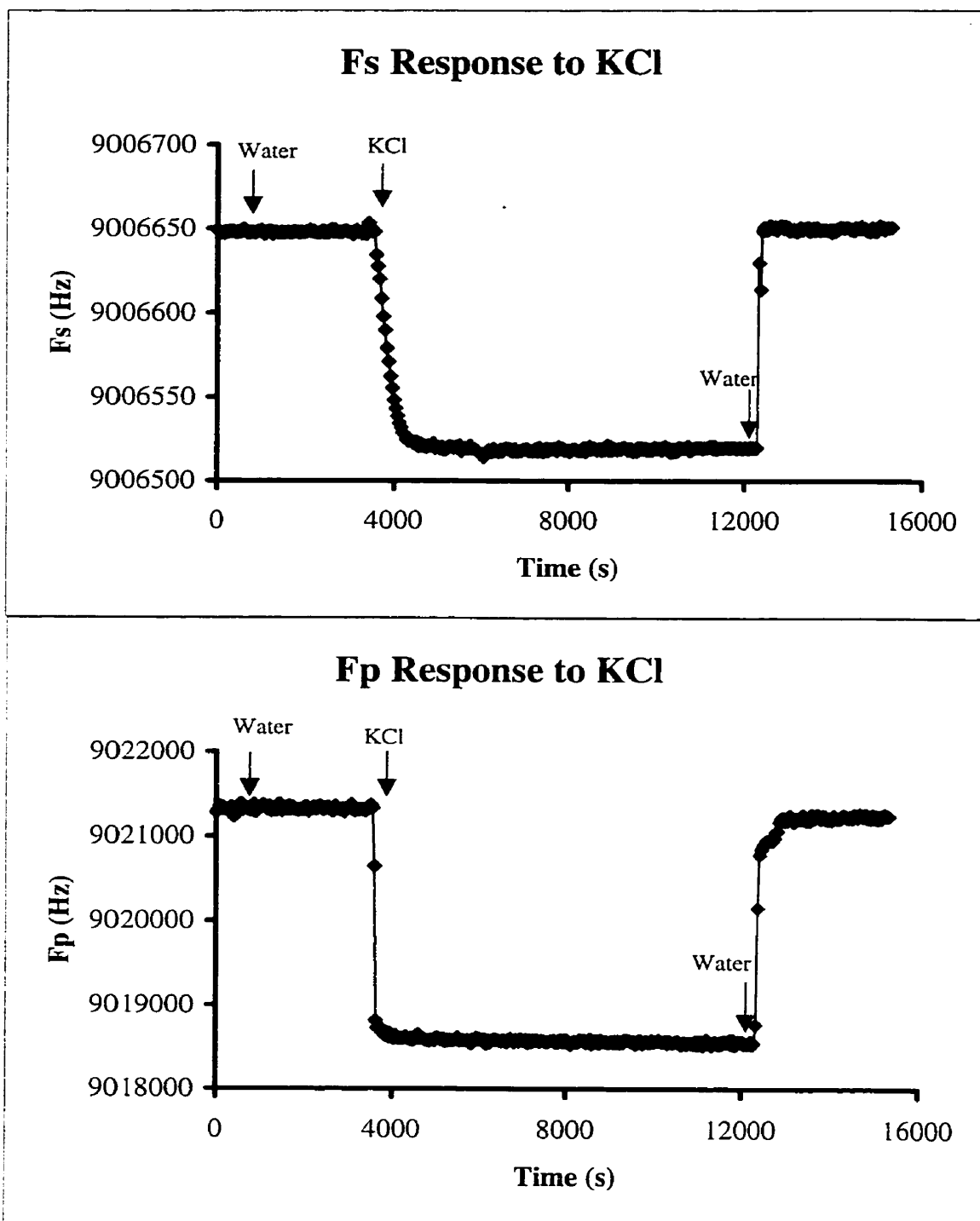
Figure 20: Responses of F_S and F_P to 0.08 M KCl solution.

94

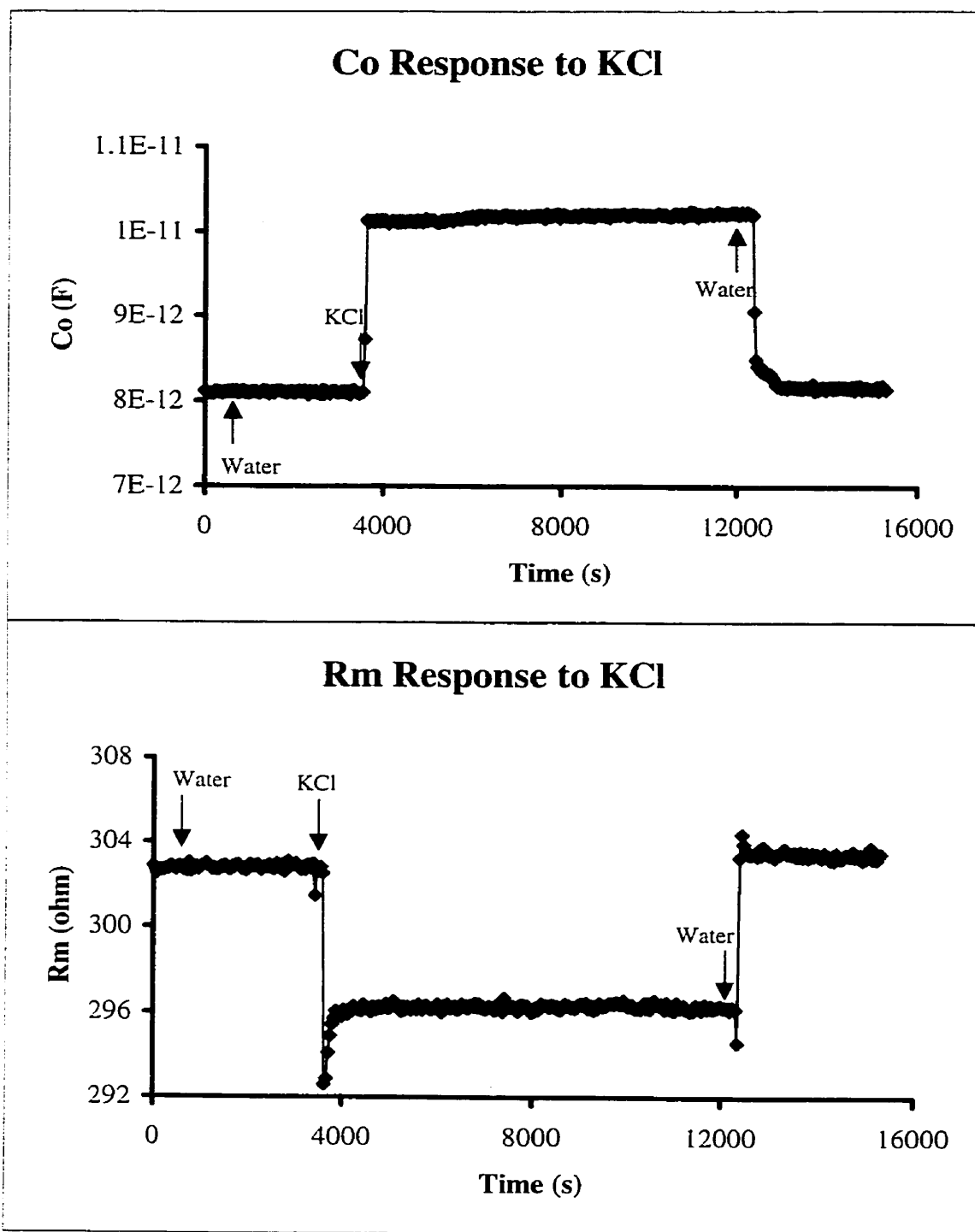# Co Response to KCl

Rm Response to KCl

Figure 21: Responses of C_0 and R_m to 0.08 M KCl solution.

95
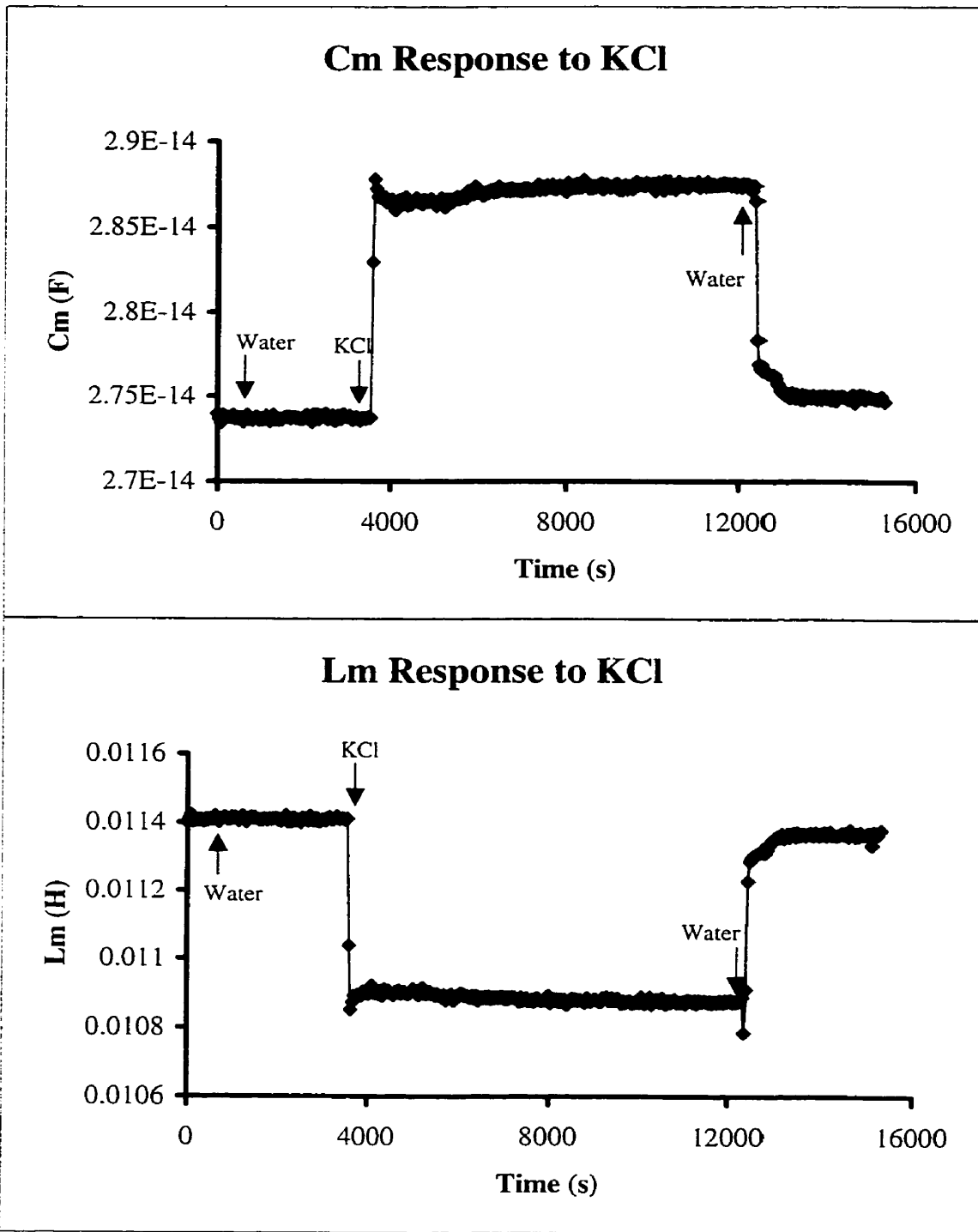
Figure 22: Responses of C_m and L_m to 0.08 M KCl solution.

Max Phase Response to KCl

Q-value Response to KCl
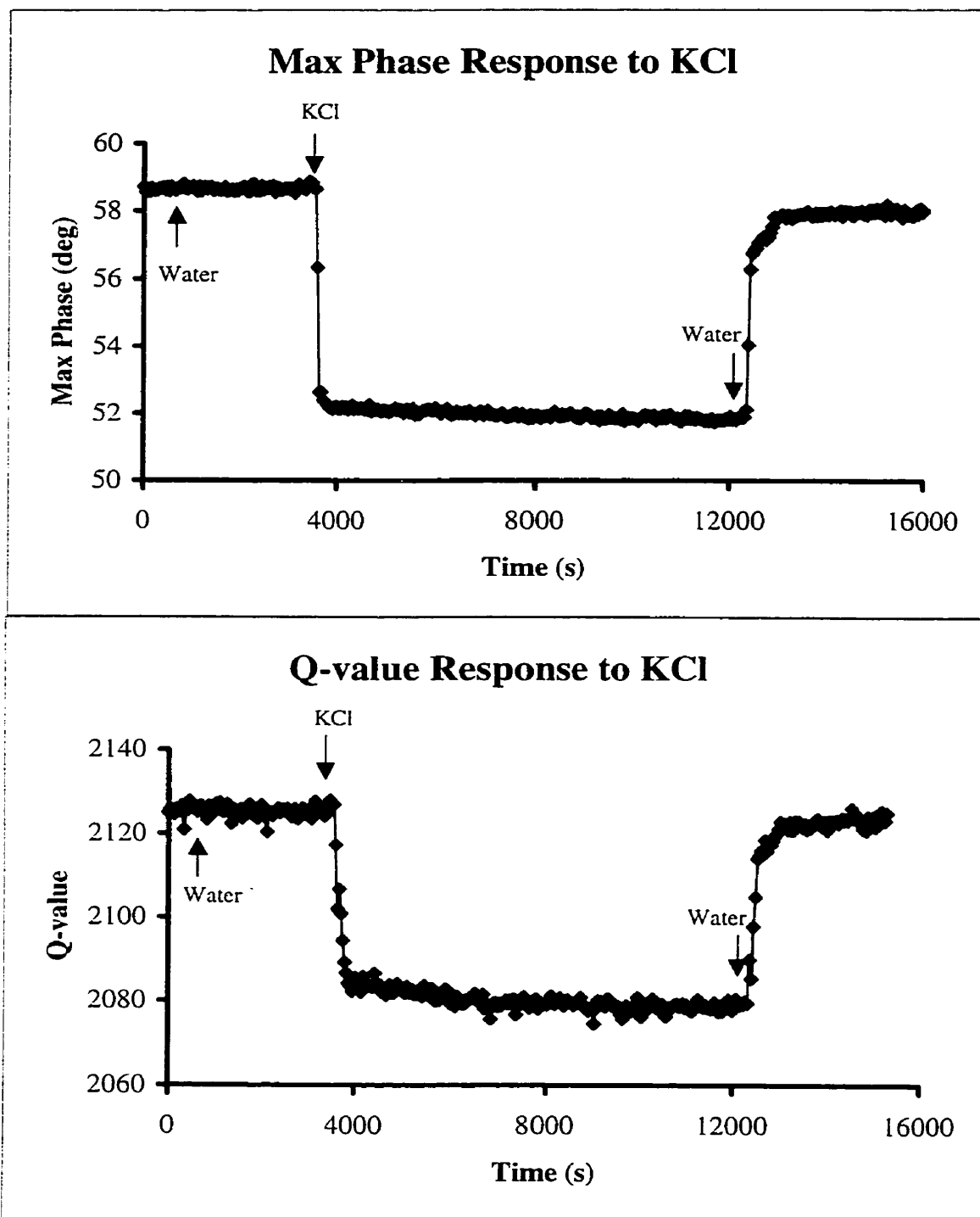
Figure 23: Responses of max phase and Q-value to 0.08 M KCl solution.

97

between this decaying electric field and the charge carriers and dipoles of the conductive solution plays a large role in the storage and dissipation of electrical energy of the system.

The charged nature of the surface also gives rise to the formation of an electrical double layer at the solid - liquid interface. Although the thickness of this double layer has been shown to be less than that of the decay length of the electric field [63], it may partially screen the electric field from penetrating into the bulk of the solution, thus effectively lowering the acoustoelectric interaction. Furthermore, the redistribution of charged particles at the solid - liquid interface may also cause a localised change in the density and viscosity of the solution near the surface, resulting in induced changes in the mass loading and viscous coupling. In addition, it has been shown that field fringing beyond the areas of the electrodes can contribute to the total change in frequency [40], particularly for the parallel resonant frequency, F_p.

Figures 24 to 27 depict changes in the responses of the same eight parameters (F_s, F_p, C_0, R_m, θ_{max}, Q-value, C_m and L_m) with respect to KCl concentrations. A common trend can be seen for all eight parameters: a relatively rapid rate of change at low concentrations followed by a slowly decreasing, almost asymptotic rate of change at higher concentrations. This oft-observed feature has been attributed to the screening action of the double layer.[37,40,64] As the ion concentration increases, the double layer tends to block more of the electric field from penetrating into the bulk solution. This is reflected in the reduction in the rates of change of the parameters with increasing concentrations. Parameters which are most sensitive to electrical perturbation, such as F_p, C_0, C_m, L_m and θ_{max}, seem to adhere to this trend more rigorously while mass and viscosity dependent parameters such as F_s, Q-value and R_m are affected but to a lesser degree. A complete treatise on the effects of conductive

98

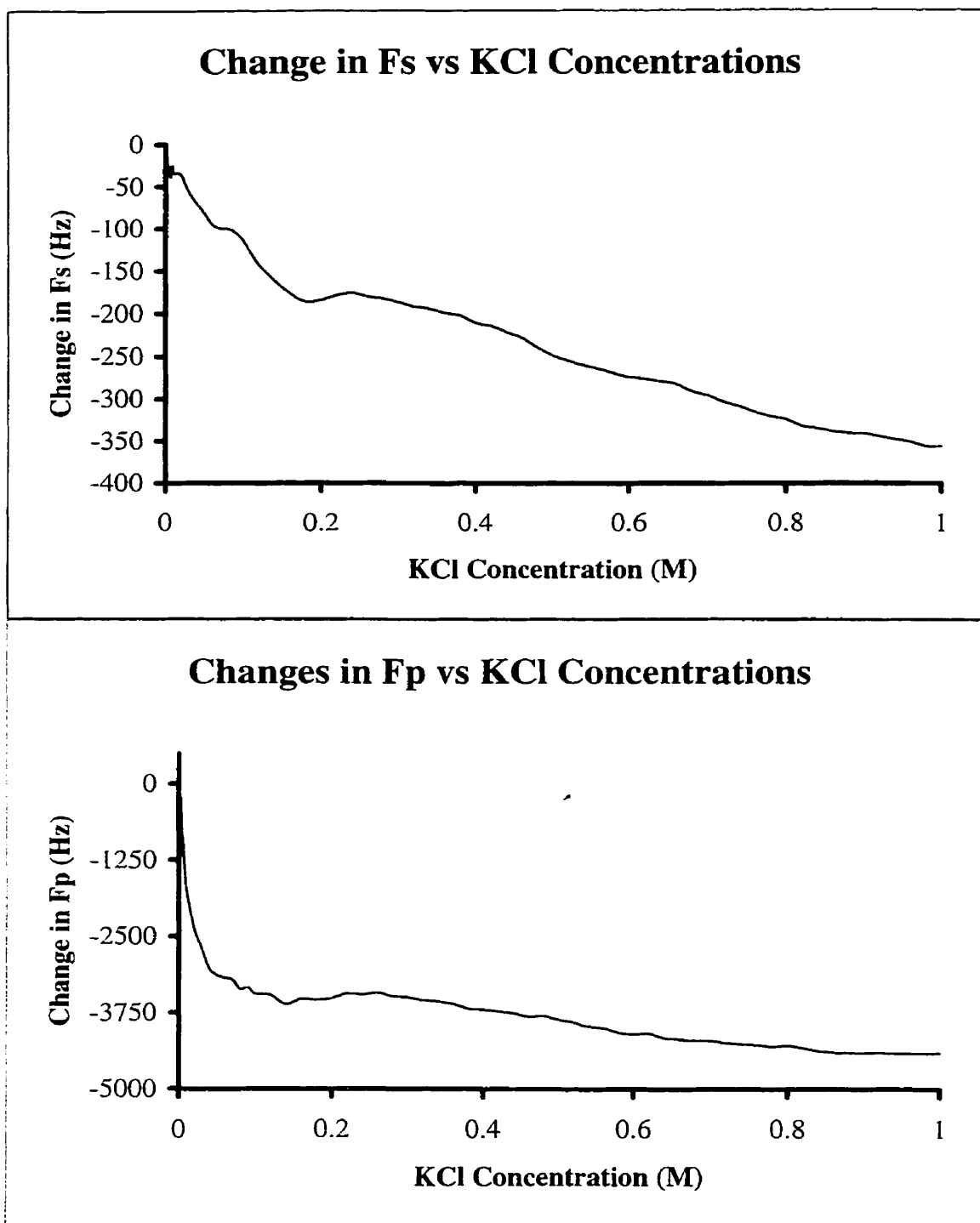Figure 24: Changes in F_S and F_P vs. KCl concentration.
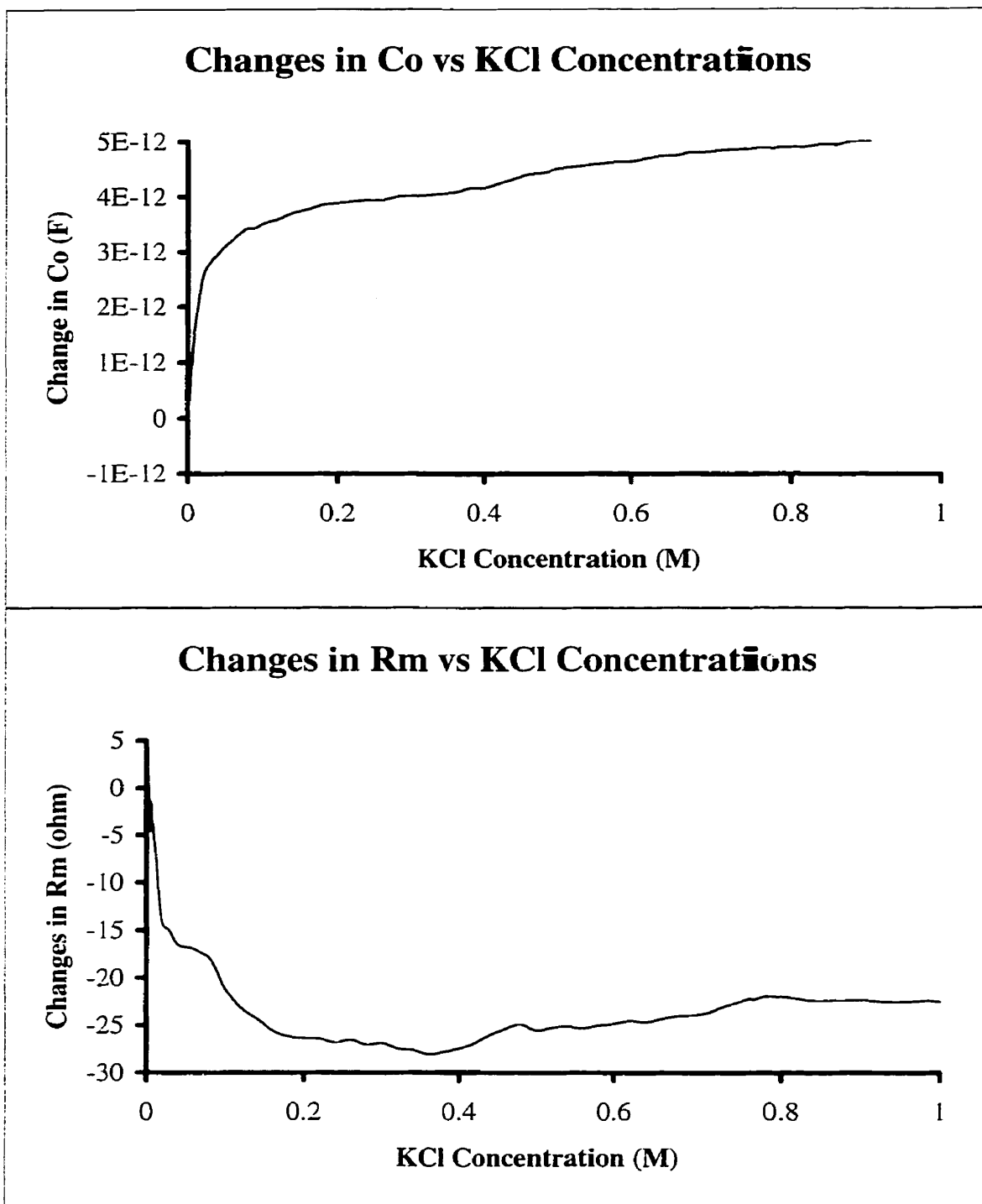
Figure 25: Changes in C_0 and R_m vs. KCl concentration.

100

Changes in Cm vs KCl Concentrations

Changes in Lm vs KCl Concentration
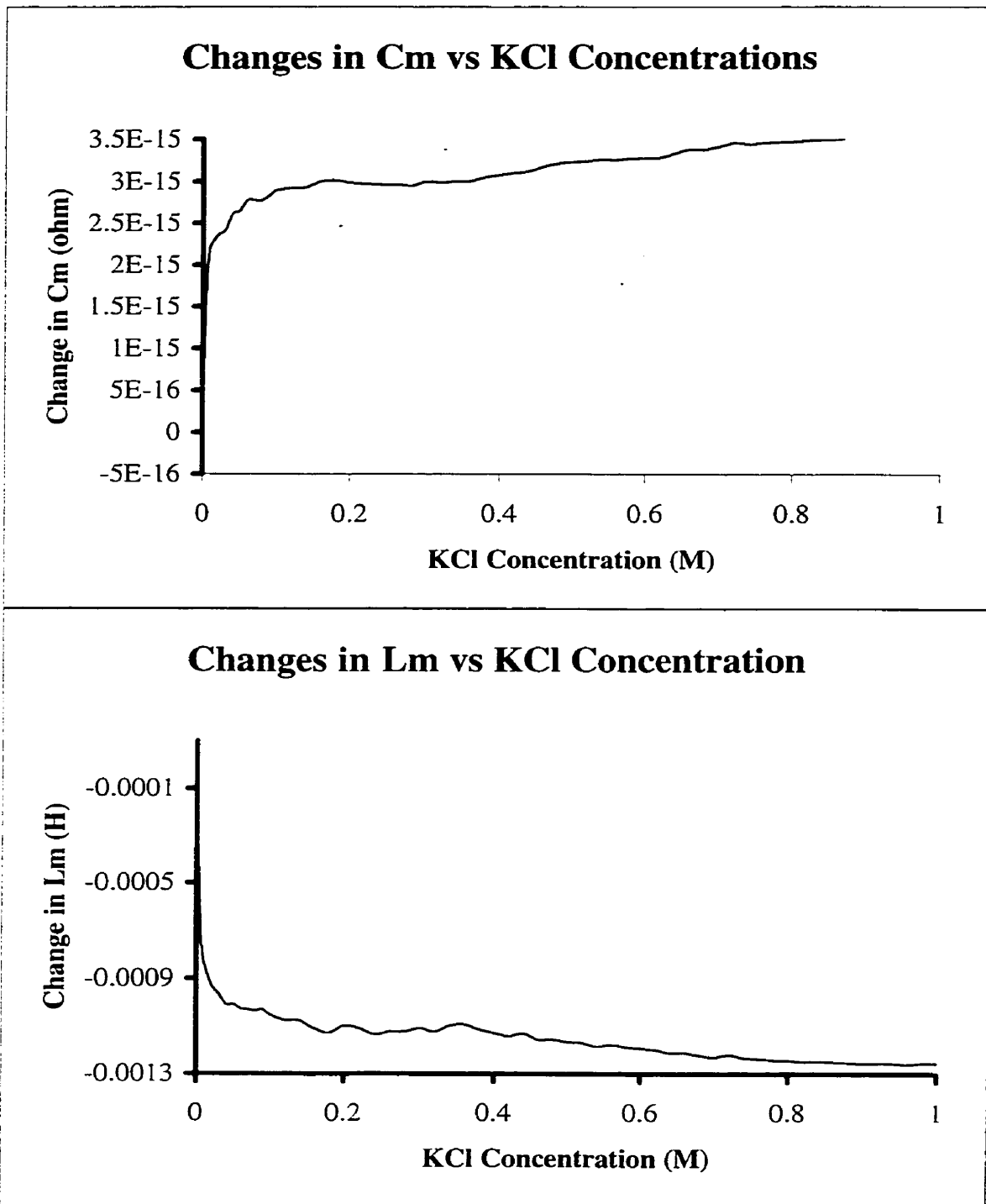
Figure 26: Changes in C_m and L_m vs. KCl concentration.

101
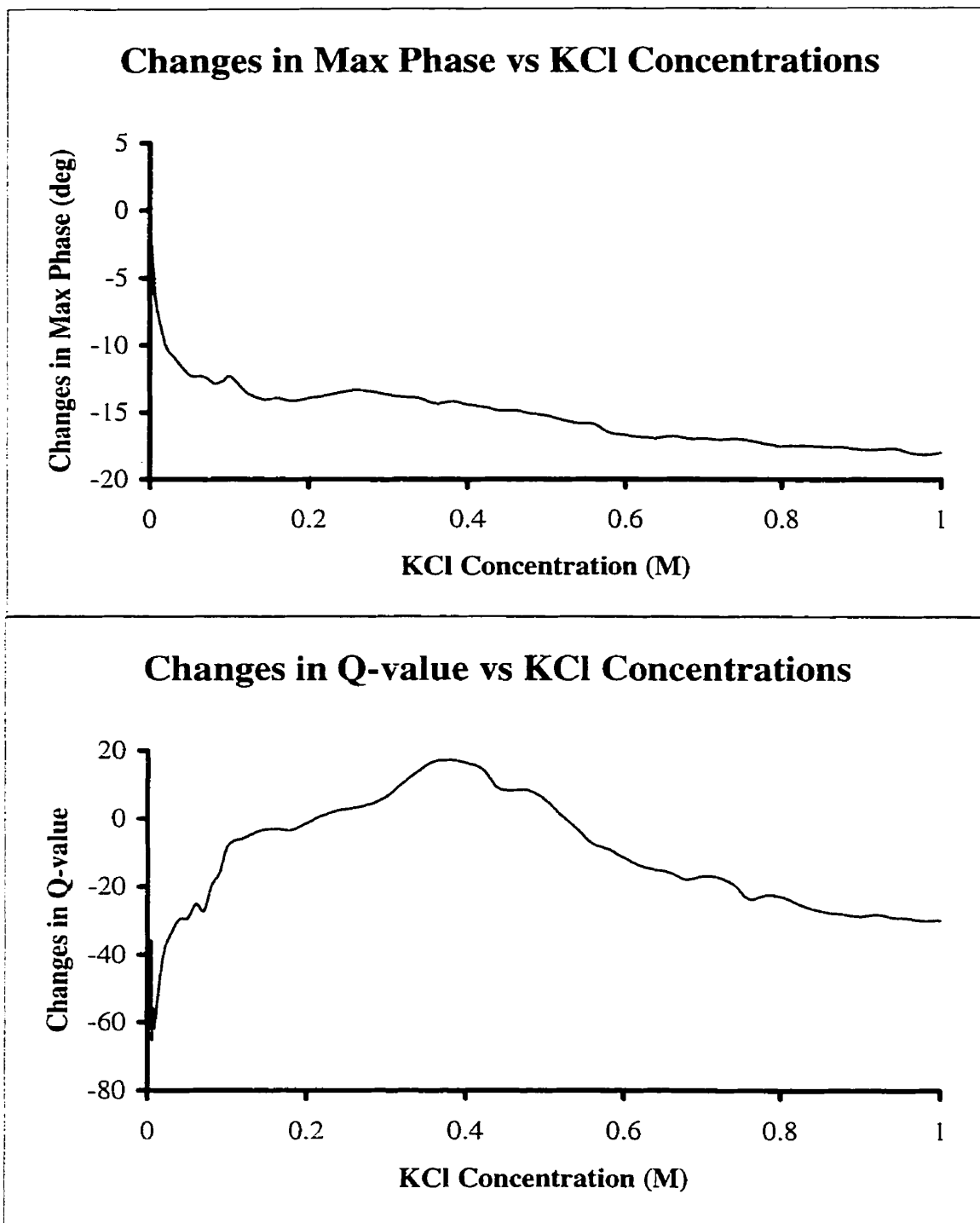
Figure 27: Changes in max. phase and Q-value vs. KCl concentration.

liquid loading on the various network analysis parameters has been presented by Yang.[86] More importantly, the non-linearities inherent in this system confirm its suitability for neural analysis.

Neural Analysis for the KCl System

While all eight parameters may be suitable as inputs, only four were chosen for the initial phase of neural computation in order to reduce the computational complexity, with the criteria for selection being relevance and independence. As can be seen from equation 27, the quality factor (Q-value) is simply a combination of the motional inductance (L_m) and the motional resistance (R_m). While the information it contains may be relevant to the system, it is already present in the other two parameters, thus Q-value may be safely eliminated. The other obvious choice for elimination is C_m since it has often been considered to be a constant under liquid loading. This is because C_m is dependent on the elastic constant, corresponding to lattice restoring forces (equation 4), and a Newtonian liquid does not possess any appreciable elasticity. Another factor which places doubt on the physical reality of any changes in C_m can be seen from its relationship with L_m. At resonance, the equivalent circuit behaves as a pure resistance since the opposite natures of the two reactive elements ($j\omega L_m$ and $1/j\omega C_m$) cancel each other out. In other words, for a specific resonant frequency, there are numerous combinations of L_m and C_m. Since the equivalent circuit parameters are derived from the regression of the measured impedance data, it is likely that changes that should be attributed to L_m may be wrongly, or disproportionally, attributed to C_m by the regression routine. It should be noted that changes in C_m are valid for systems in which a viscoelastic material is adsorbed onto the surface of the sensor as in studies involving polymer films. However, this is

103

clearly not the case here thus C_m can be removed as a possible input.

The elimination of the next two parameters as inputs was slightly more difficult. A neuralnet-based pruning process was required since there was no justification based on theoretical considerations alone. A reference network was trained on the KCl data, with all six remaining parameters as inputs and the concentration as the output. The training root-mean-square error (RMS) for this reference network was used as the benchmark error, against which subsequent architectures were compared. Next, one parameter was removed and the resulting network was trained on the appropriate training set. This exercise was repeated , each time removing one input parameter. The two parameters, whose removal caused the least and the second least adverse changes in the training RMS, as compared to the benchmark RMS, were eliminated from the input pool. Since they contributed the least to the goodness-of-fit of the neuralnet, these two parameters can be considered to be less sensitive and /or less independent than the others. The results of this process are depicted in table 3. It can be seen from the RMS results that L_m and θ_{max} are the two least effective inputs, *i.e.* they contribute less to the regression than the other four parameters, and can thus be eliminated. This, however, does not disqualify them from future calculations for which more inputs are desired. It is instructive to take stock of the four input variables deemed to best describe the electrolytic system. In F_s, the effect of mass loading on the sensor is well represented while R_m will give a clear indication of the viscous coupling affecting the energy dissipation of the system. The static capacitance, C_0, will describe the electrical properties of the conductive solutions while F_p will be affected by the capacitive loading, generated by the electrolyte, as well any mass loading, with the former dominating.

Once the optimal set of inputs has been determined, neural computation was

ARCHITECTURE	PRUNED PARAMETER	RMS
Reference (6, 6, 1)	None	0.00219
P1 (5, 6, 1)	F_S	0.00681
P2 (5, 6, 1)	F_P	0.00594
P3 (5, 6, 1)	C_0	0.00517
P4 (5, 6, 1)	R_m	0.00778
P5 (5, 6, 1)	L_m	0.00279
P (5, 6, 1)	θ_{max}	0.00426

Table 3: Determination of optimal input parameters.

performed on both software platforms, using the KCl data sets. For the "unadulterated" system, training and testing data sets were constituted of 100 and 50 instances, respectively, covering the range of concentrations between 0.001 M and 1M. The concentrations of the "unknown" samples in the validation set were also in this range. A typical NeuralWare network, with a 4,6,1 architecture: 4 input nodes, 6 processing nodes in one hidden layer and 1 output node, is shown in figure 17. A bias is used to ensure the nodal outputs are in the optimal range and the window in the top left hand corner tracks the decreasing RMS error. It can be seen from the RMS graph that for a typical viable network, the RMS decreases in a near exponential fashion.

A typical output of a trained ANNAT network is depicted in table 4. This represents a 4,5,1 network with 30 weights reaching convergence in 11000+ iterations. Each of the weights can be related back to a specific connection in the network. The relevance of this is that it allows each of the weights to be traced back to a specific input, an important factor for the subsequent weight adjustment process. The sum-of-squares error (SSQ) shown was converted to RMS for meaningful comparison with the NeuralWare platform. The optimal architecture of any network is usually derived through a trial and error process. Figure 28 illustrates the relationship between training RMS error and the number of hidden nodes for a network with one hidden layer. The RMS shows an initial rapid improvement with increasing number of hidden nodes, followed by a slower and decreasing phase of improvement. After a certain number, the slight improvements no longer justify the extra computing time and effort, and the network is set. The optimised network is then subjected to the test data set to determine the testing RMS error. For a network to be viable, i.e. one which is able to make generalisations, the testing RMS error should be similar to the training RMS error. At the very least, it should be of the

106

Number of input and output nodes: 4   1
Number of hidden layers: 1
Number of nodes in each hidden layer: 5
Activation function for output nodes: 1
Activation function for hidden nodes: 1
Gain term - beta: 1.0000
File with input patterns and outputs: kcl.dat
Print option (0/1/2):   0
Maximum number of iterations: 20000
Analytical/numerical derivatives <A>:   A
Limit for initial weight-guesses: .00000E+00
Seed: 9823

Number of patterns in the input file: 100
The weights W >>> 1 to 30 :

7.6818	-.31707E-01	.69635E-02	-.14397	3.6978
-48.077	-.12729E-01	-.42298E-02	.23473	7.0486
7.304	.22442E-01	-.13044E-02	.86784E-01	-3.7655
1096.1	-1097.0	-1.0025	-4.8776	1.0893
-25.790	.90235E-01	-6.0004	-.34710E-0	291.772
3.1639	-66.097	.36891E-01	.10492	-69.200

SSQ : .44755E-03

Number of iterations : 11124

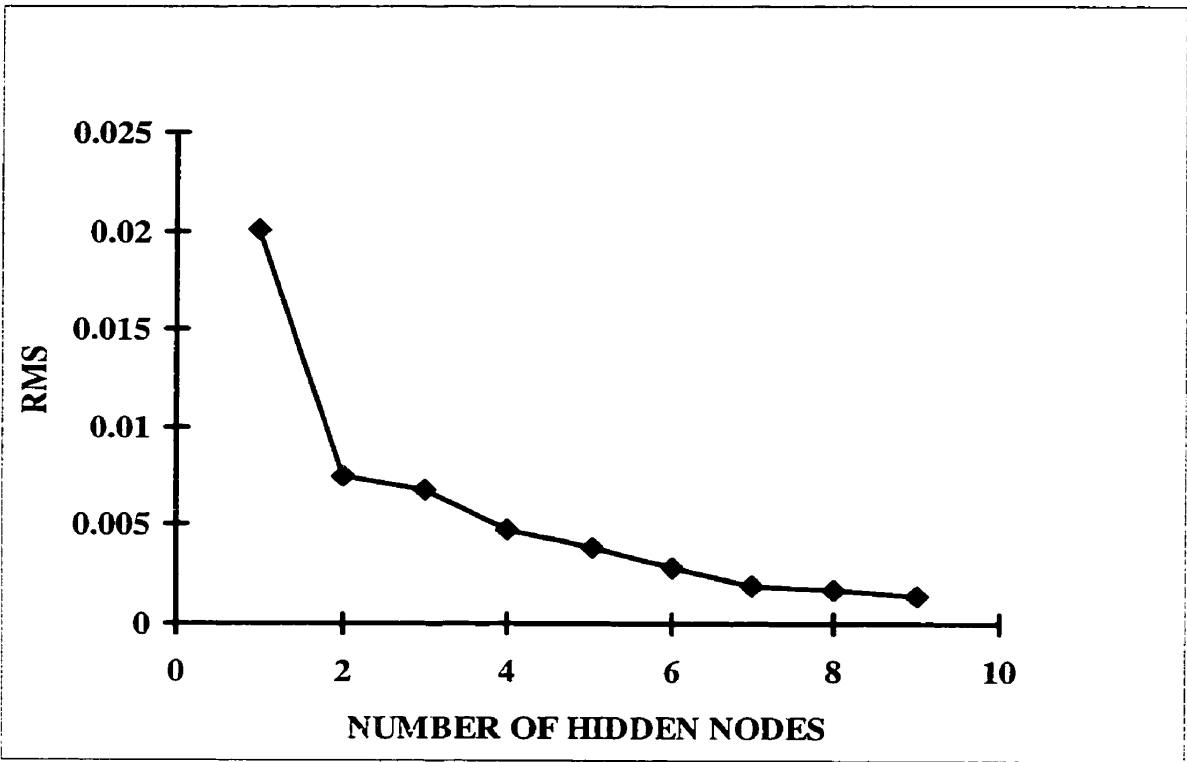Table 4: Typical output for ANNAT training run.

107

Figure 28. Training RMS vs. Number of Hidden Nodes

same magnitude as that of the training set. A network which trains well but does not test well is likely to be an overtrained network. If testing proves the network to be unsuitable then a different architecture must be chosen, usually one with a smaller number of hidden nodes. The new network would then be tested to ensure viability, prior to usage.

The results for the neural calculation on the "unadulterated" KCl data sets are depicted in table 5. Using the ANNAT package, the best architecture obtained is a 4,7,1 net with no further significant improvement after 10000+ iterations. The RMS for the test set can be seen to be of the same magnitude as that of the training set, validating the network. For the commercial package from NeuralWare, the best architecture is 4,9,1 and the number of iterations required is approximately doubled. The RMS for the training and testing sets are also of similar magnitudes, indicating a useful architecture. Slightly lower training RMS was observed for NeuralWorks but the test RMS for this program was actually slightly higher than that of ANNAT. The discrepancy is not large but it could be argued that the larger NeuralWorks network, although it may have trained better, may not generalise as well as the smaller ANNAT architecture.

Thirty five solutions of various concentrations in the 10^{-3} M to 1 M range were culled from the test set to become the unadulterated validation set. The response sets for these "unknowns", comprising of F_S, F_P, C_0 and R_m, were used to test the predictive ability of the trained networks. Using ANNAT, these predictions resulted in errors in the range of 3% - 7%, compared to the true values (table 5). This can be considered to be acceptable given the relatively low precision of neural analysis, compared to other regression methods, and the large range of concentrations under investigation. A network optimised for a smaller range of concentrations will

109

PARAMETERS	ANNAT	NeuralWorks
Best architecture	4,7,1	4,9,1
Number of iterations	10000 +	20000 +
RMS (training)	0.00142	0.00129
RMS (test)	0.00407	0.00445
Predictive error (unadulterated)	3%-7%	2%-7%
Predictive error (thiol, unadjusted)	8%-14%	7%-15%
Optimal weight reduction	18%	21%
Predictive error (thiol, adjusted)	3%-8%	3%-9%

Table 5: Comparison of results from ANNAT and NeuralWare.

undoubtedly yield a corresponding reduction in the error range. The predictive ability of the commercial package, with respect to the unadulterated unknown data sets, is slightly better than that of ANNAT (table 5). This is somewhat surprising in light of the higher test RMS for NeuralWorks. The differences are small enough that they can be attributed to the culling process, *i.e.* the samples chosen responded better to the NeuralWare package, despite the fact that care was taken to select a validation set which spans the range of concentrations under study. For both systems, no clear trend in the % error can be observed, with respect to concentration, but the highest % errors seem to occur for lower concentrations (figure 29).

Neural Analysis for the Thiol-Adulterated KCl System

Neural network analysis has been shown to be effective in handling ideal data sets but this alone does not justify the time and effort spent generating the vast amount of data and creating the appropriate architecture. The value of neural computation lies in its ability to classify less-than-ideal data sets, using the same network trained and tested on the exemplary data sets. This next step involved the exposure of the trained networks to data generated by samples that have been adulterated with an interferent, ethanethiol (CH_3CH_2SH), hypothesised to affect only one of the four input variables. This compound was chosen specifically so that the thiol functional group will interact with the gold electrodes of the TSM device to generate a mass loading effect and induce a change in the series resonant frequency, F_S. Although a larger molecule will cause a larger deviation in F_S, ethanethiol was chosen in view of its solubility in an aqueous system. The shorter chain length also precludes formation of a well-packed monolayer. This, combined with a purposely low concentration of thiol
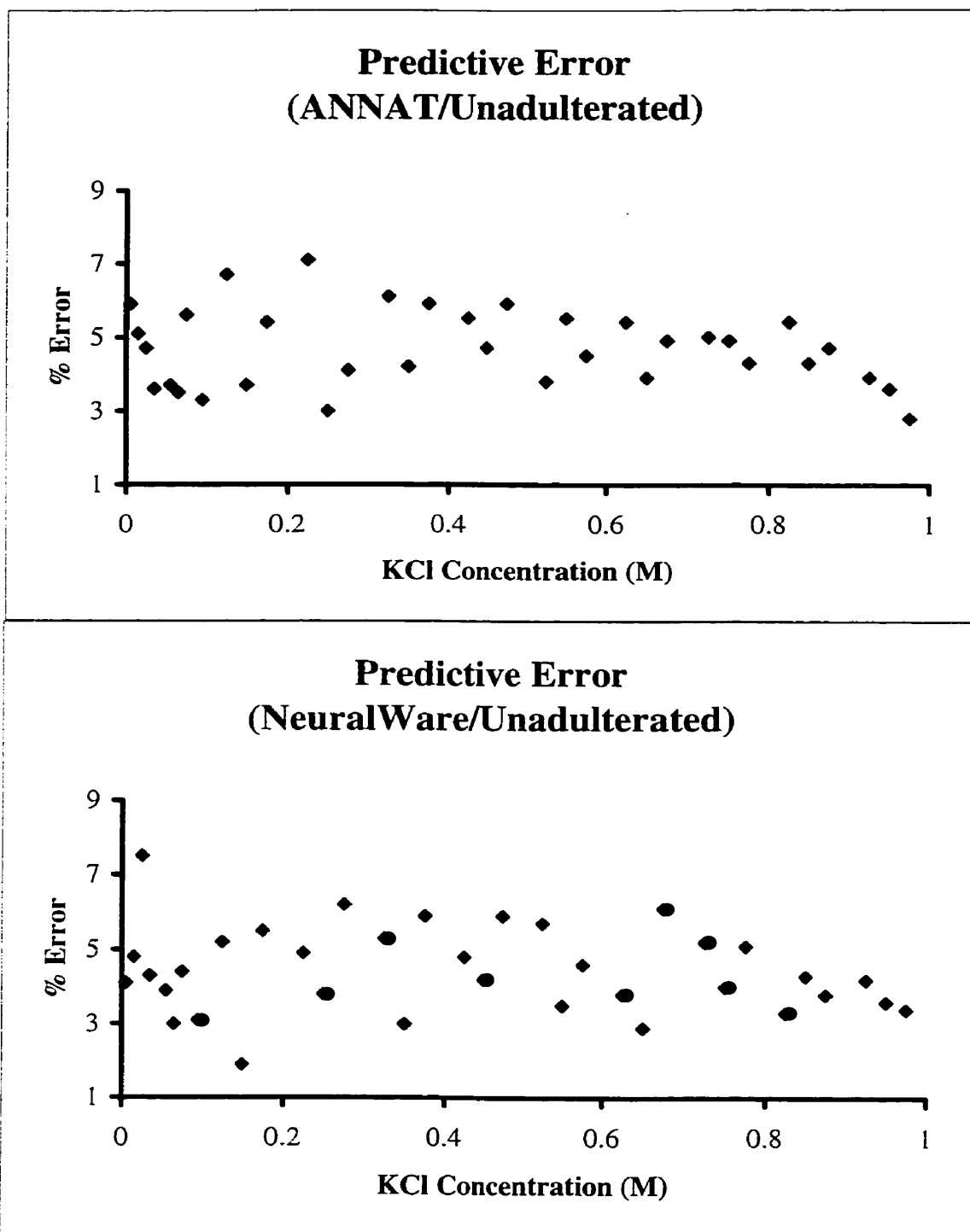
111

Predictive Error
(ANNAT/Unadulterated)

Predictive Error
(NeuralWare/Unadulterated)

Figure 29: Error distribution of unadulterated unknown.

112

(<mmol), ensures that changes in the viscosity, density and electrical conductivity of the sample are kept to a minimum so that the change in F_S can be attributed mainly to mass loading. An important corollary effect is that the other three parameters, R_m, C_0 and F_P should remain relatively unchanged in order to prevent too large of a perturbation of the network. In an effort to ensure meaningful comparison, the thiol was introduced into samples of the same KCl concentrations as those in the unadulterated validation set.

The effects of the contaminant on the responses of the TSM sensor are illustrated in figures 30 and 31. As expected, a significant relative change in the series resonant frequency can be observed upon addition of the thiol interferent. A quick drop followed by a more gradual downward trend concur with the accepted adsorption profile for the Au-SH system.[86] The adsorption process seems to be permanent, as indicated by the lowered frequency even after the flow was switched back to pure water. However, the magnitude of change does not agree with the classical Sauerbrey mass loading effect, even if a dense packing of thiol can be assumed. The increased response can be attributed to changes in density, viscosity and electrical properties induced by the addition of the thiol. These changes are confirmed upon examination of the other network analysis parameters as they all registered small but significant changes. For all three variables, the switch back to pure water restored the responses to near starting levels. The changes induced by the thiol were too small, with respect to the noise inherent in the signals, to be verified as transient or permanent.

The set of adulterated "unknown" data was processed by the networks, trained on the unadulterated data sets, to determine the robustness of the neural computation. As can be seen from table 5, the predictive errors for the adulterated "unknowns" were significantly higher than those for the unadulterated "unknowns", between 8% - 14%
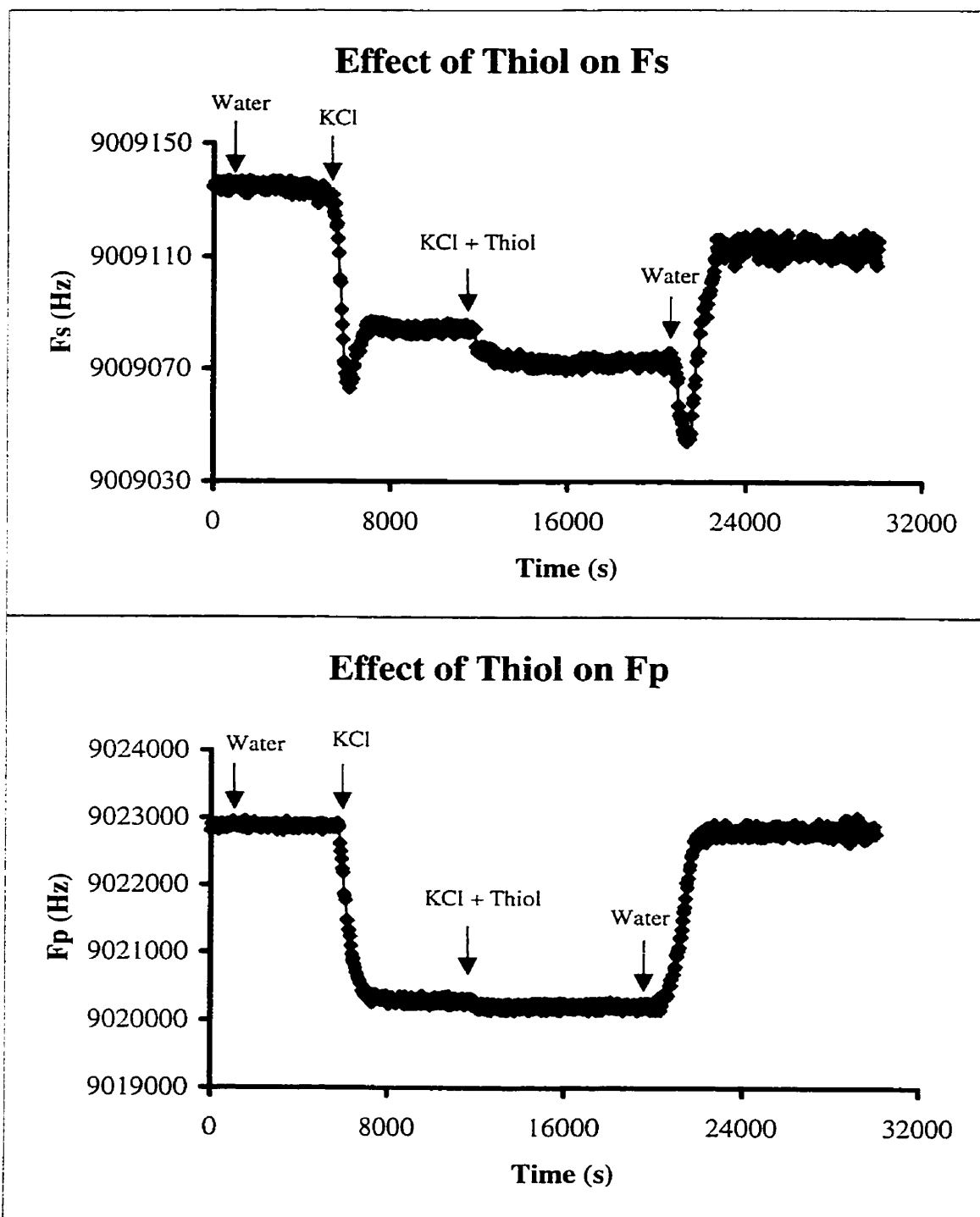
113
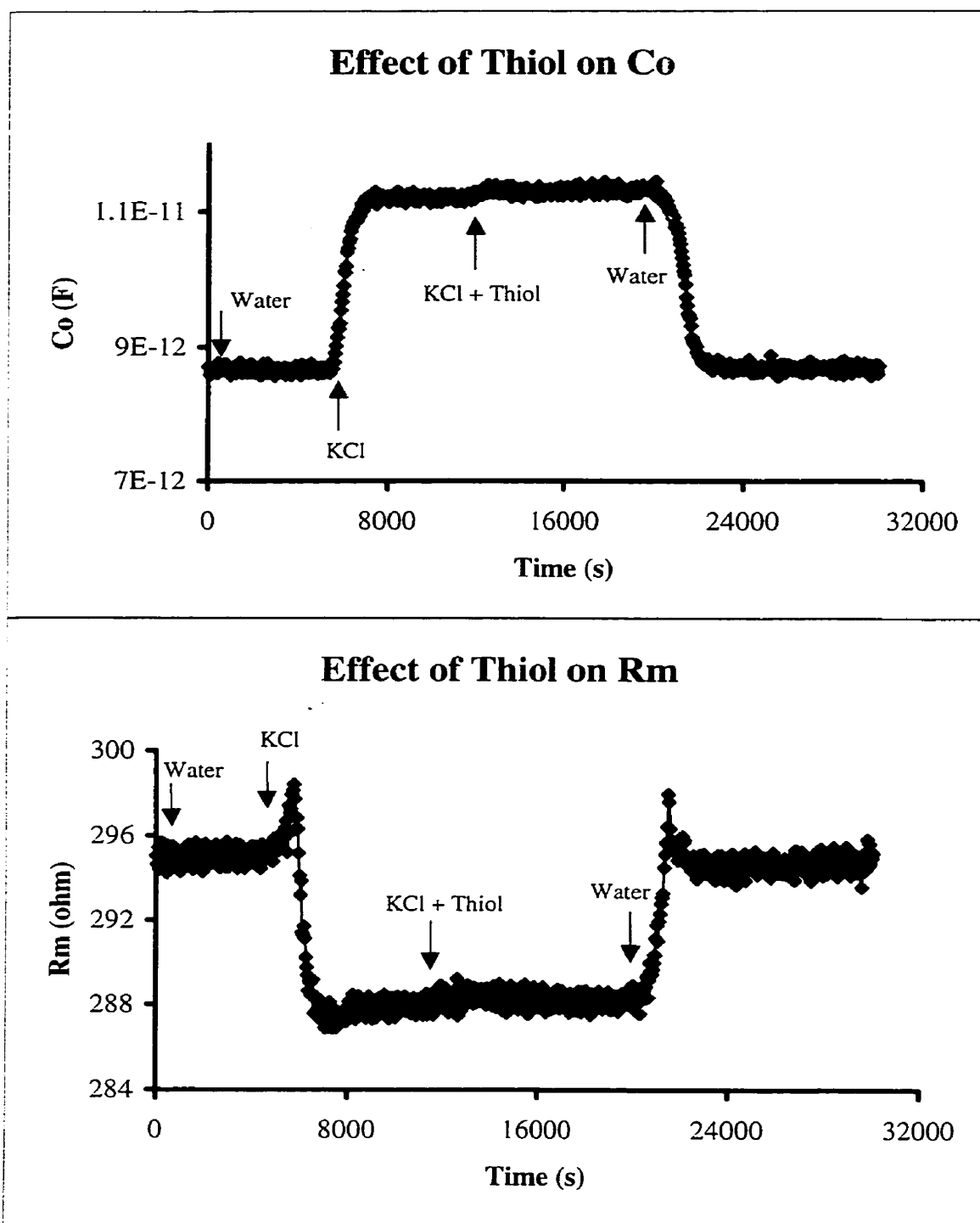
Figure 30: Effect of thiol on F_S and F_P.

Figure 31: Effect of thiol on C_0 and R_m.

for ANNAT and 7-15% for NeuralWorks. These results established the neural networks to be fairly robust, but it is obvious that a network trained using unadulterated I/O data cannot be used to predict adulterated samples accurately. Once again, no clear trend for the % error was observed, with respect to concentrations (figures 32 and 33), but the predictive errors for NeuralWorks were now slightly higher than those for ANNAT. This seems to indicate that while the larger network may generalise an ideal data set better, it does not possess the same robustness for tainted data.

Weight Adjustment - As hypothesised, the thiol-induced changes in F_S were found to be significantly larger for the adulterated samples, compared to the changes in the other three parameters. It is clear that the degradation in the performance of the trained networks can be attributed mainly to the perturbation of the series resonant frequency. Knowing this, several possible procedures can be considered to rectify this situation. The most effective solution would be the development of a new set of networks, trained and tested on adulterated data sets. However, this would involve the generation and network analyses of over one hundred new KCl solutions, with the appropriate contamination. In addition the whole training, testing and validation process must be repeated for the new set of data. A less drastic measure would be the elimination of F_S as an input parameter from the original data sets. This may prove to be useful but it would still involve the repetition of the training and testing processes to find the optimal architecture for a three-input system. Even then, there is no guarantee that the networks trained and tested on three-input, unadulterated data sets will be effective for the adulterated "unknowns". An analogous solution is to increase the number of input variables, making use of the other available parameters, in order
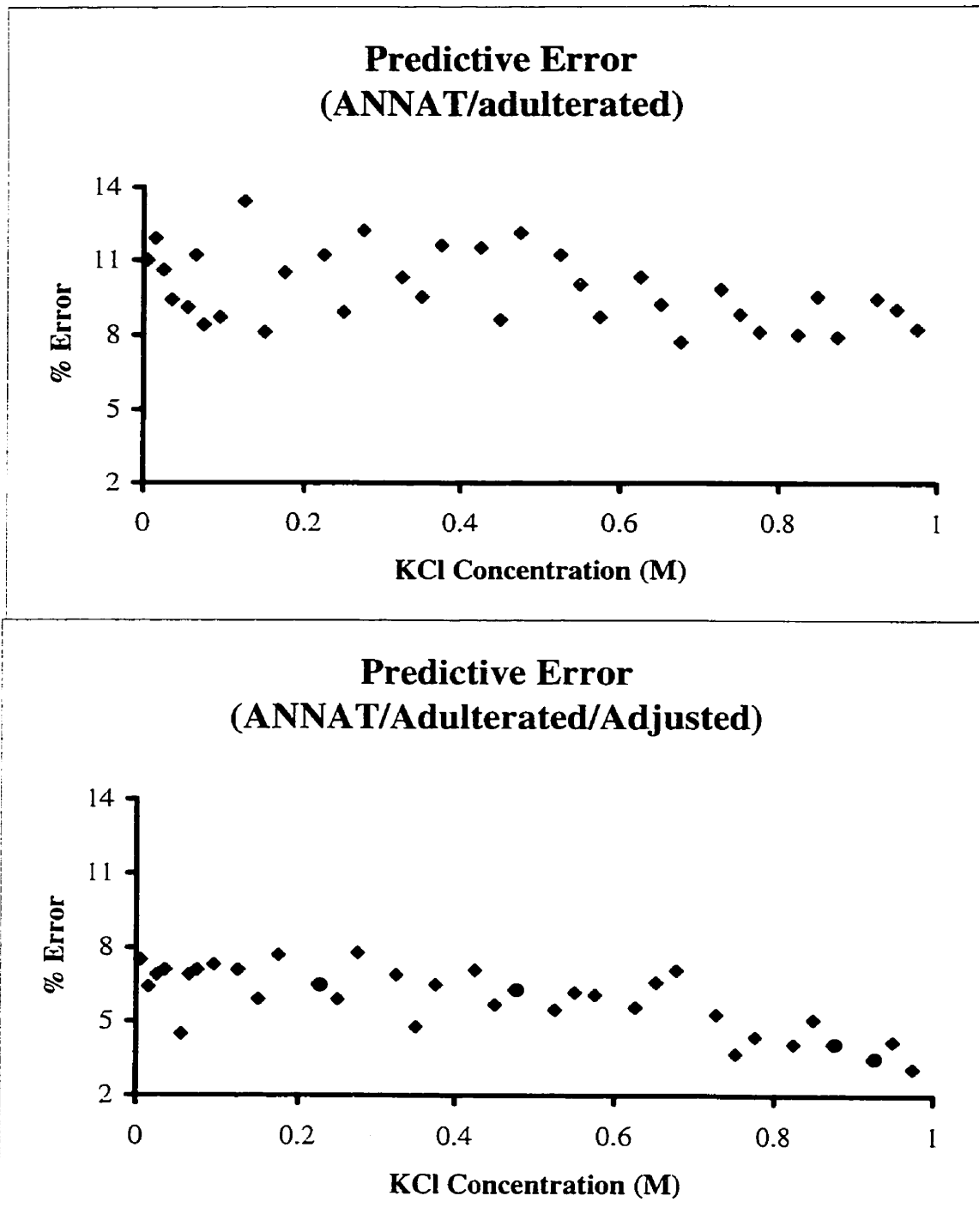
116

Figure 32: Error distribution for ANNAT results.
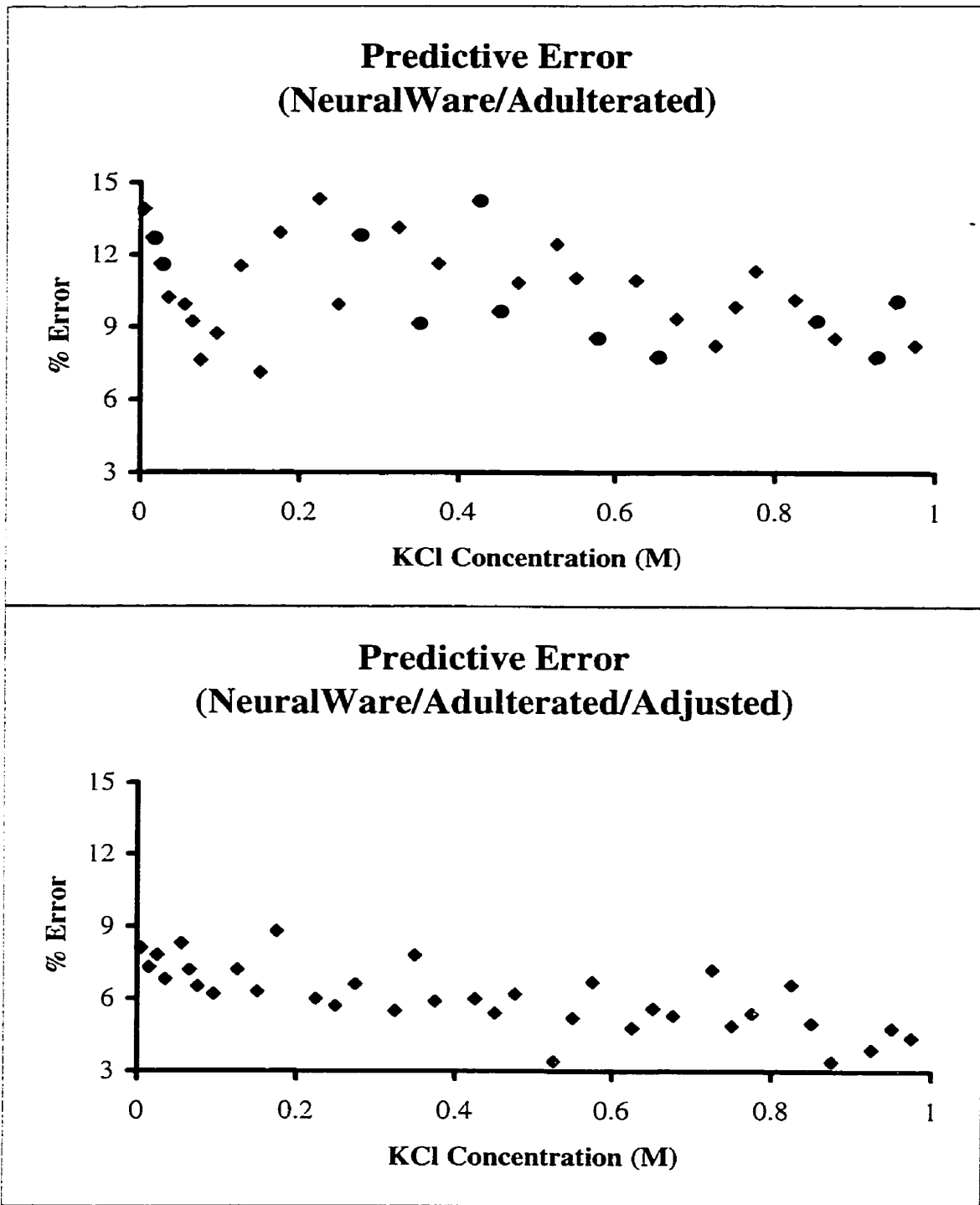
Figure 33: Error distribution for NeuralWare results.

to reduce the effect F_S has on the trained networks. This may indeed increase the robustness of the network but once again it involves time consuming training and testing procedures which are certain to be more intensive due to the extra complexity introduced by the larger data set.

An effective yet facile answer may lie in the manipulation of the existing networks. For any trained network, each of the weights can be related back to a specific connection in the neuralnet. It was hypothesised that manipulation of the weights, which are associated with the F_S - to - hidden layer connections, may provide a possible modulation mechanism of the F_S input and this, in turn, may temper the effect of the adulterant. It was discovered that a reduction in the magnitudes of these weights resulted in an improvement in prediction accuracy for the adulterated samples for both software packages. A process was developed in which the F_S - related weights were reduced incrementally, followed by testing with the whole adulterated data set. The RMS error of each weight-reduction increment was monitored until the optimal RMS error was achieved. The adjusted network would then be used to predict the adulterated "unknowns" individually. Optimal results were obtained with reductions of approximately 20% in the magnitudes of the affected weights; 18% and 23% for ANNAT and NeuralWorks, respectively (table 5). The similarity in the magnitudes of the adjustment indicates that this is a systematic phenomenon and not a random one. The predictive errors achieved for the adjusted networks are close to those for the unadulterated samples: 3% - 8% for ANNAT and 3% - 9% for NeuralWorks, respectively (table 5). The small decrease in efficiency may be attributed to the fact that this procedure does not take into account the thiol - induced perturbation on the other three input variables. The distribution of the % error, with respect to concentration, are illustrated in figures 32 and 33. For both software

119

platforms, in addition to the reduction in magnitude, a decrease in the variance of the data points is also observed. This may be a direct consequence of the weight-adjustment procedure as the detrimental effect of the thiol on the predictive ability of the neuralnet may have been negated by the adjustment process.

The manipulation of the weights associated with the affected parameter seemed to have lessened the deleterious effect of the interferent, without having to resort to the construction of new networks. In effect, the existing networks have been calibrated for the presence of the thiol. This procedure can be regarded to be roughly analogous to the temperature calibration process for a pH meter. While the results are encouraging, more work needs to be done before this weight-adjustment procedure is better understood and validated. The introduction of a second and further target interferent will help to elucidate the weight-adjustment process.

Neural Analysis for the Glycerol-Adulterated KCl System

An analogous experiment was performed in which the KCl system is subjected to a different interferent, followed by the appropriate neural computation, in an effort to clarify and validate the weight - adjustment procedure developed on the thiol system. The same criterion, used in the former system, is applicable here; that is the chosen interferent should, preferably, affect only one of the four input parameters. Of the remaining input variables, both F_p and C_0 will respond to a perturbation in the electrical property of the analyte. The default target is, thus, the motional resistance, R_m. Since the motional resistance tracks the energy dissipation, via viscous coupling with the contacting solution, of the TSM device, the interferent should be chosen to affect the viscosity of the KCl system, without disturbing its electrical properties.

120

While it is clear that a change in viscosity will also have an effect on the series resonant frequency [33,34,35], a judicious dispensation of the interferent may minimise this effect. The obvious candidate, utilised in numerous viscosity-perturbation investigations [62,87], is glycerol.

The effect of glycerol on the responses of the TSM sensor are illustrated in figures 34 and 35. As expected, a significant change in the motional resistance can be observed upon addition of the glycerol interferent. An increase in the R_m value indicates a corresponding increase in the energy dissipation, in good agreement with the higher viscosity of the solution. More importantly, the capacitive related responses, C_0 and F_P, did not show any appreciable change, denoting an invariance in the electrical properties of the system. Unfortunately, even though the concentration of glycerol was kept deliberately low, a change is the series resonant frequency was observed. As undesirable as it was, this is a logical response to the increase in density, thus a change in viscous loading, brought about by the addition of the glycerol. The higher density is also reflected in a slight decrease in F_P, confirming the fact that although the parallel resonant frequency is sensitive to both mass loading and capacitive loading, the latter is the dominating factor. Figure 36 depicts the changes induced by glycerol on the motional inductance (L_m) and the motional capacitance (C_m). Both parameters seem to be immune to the presence of glycerol, making them strong candidates for inclusion in any future glycerol-based investigation.

The results of neural computation on the glycerol-contaminated solutions are shown in table 6. Since previous studies did not reveal any significant difference between the performances of the two software platforms, a decision was taken to eliminate ANNAT from subsequent investigations. The main reason for this decision
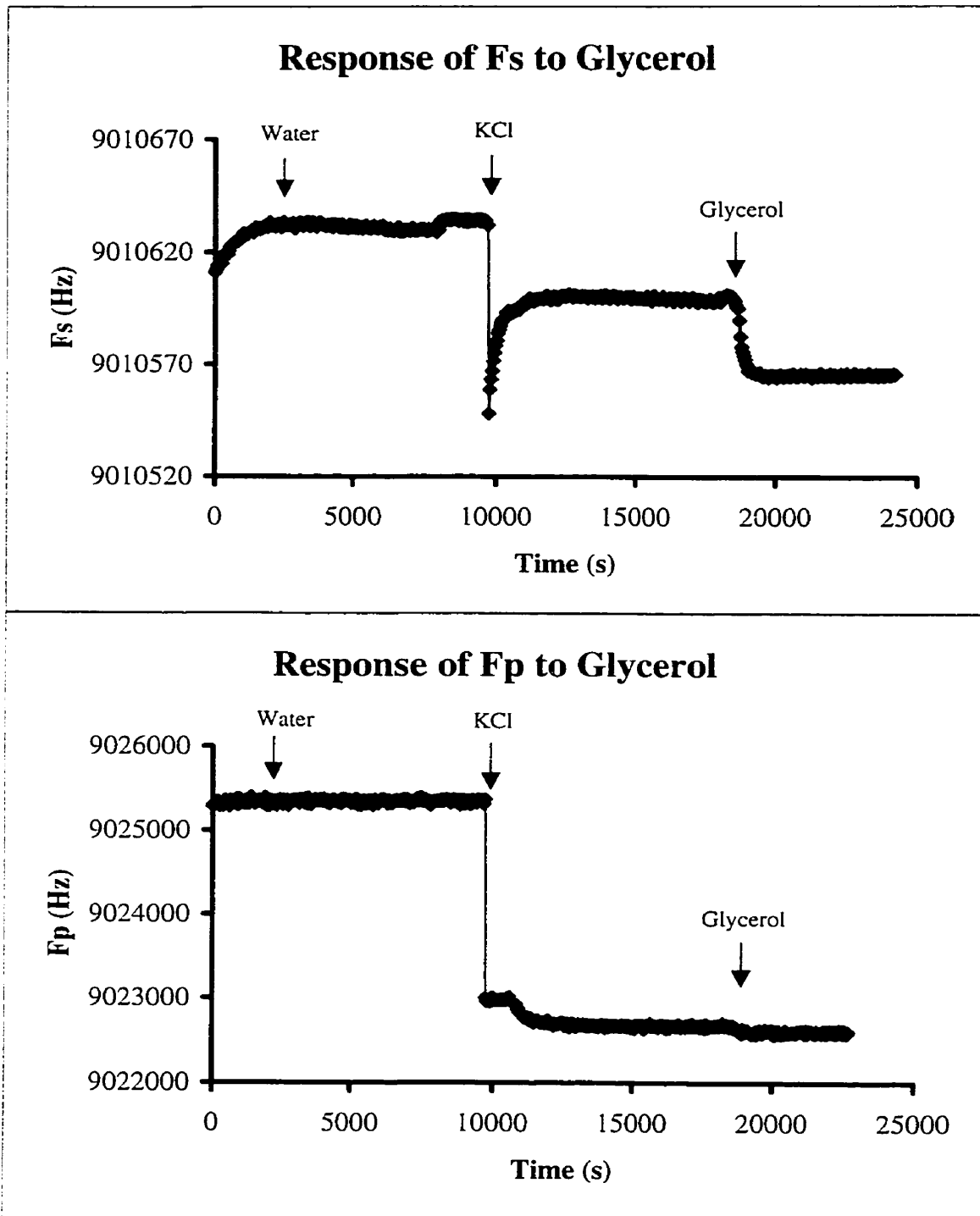
Figure 34: Responses of F$_S$ and F$_P$ to glycerol (0.04 M KCl solution).
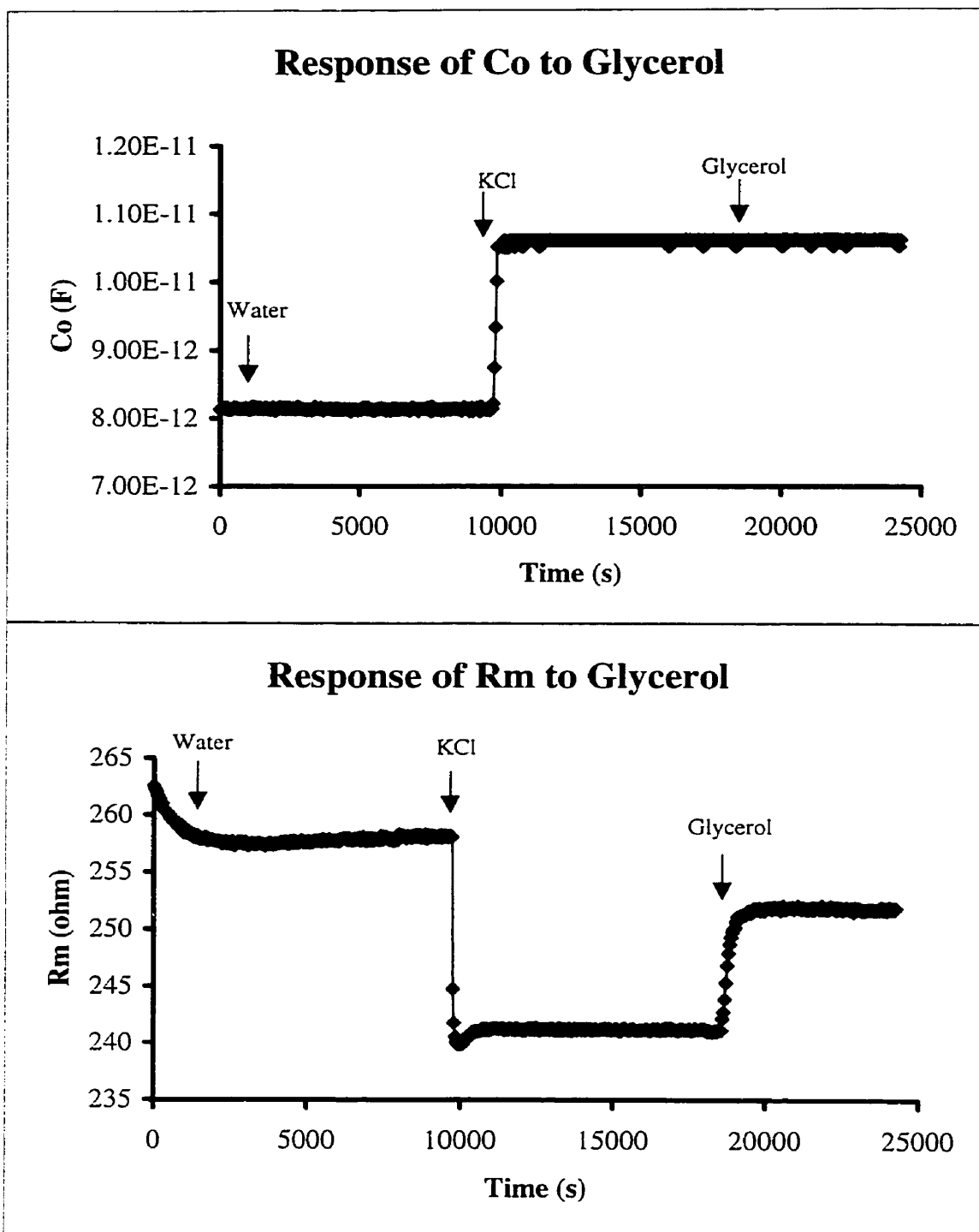
Figure 35: Responses of C_0 and R_m to glycerol (0.04 M KCl solution).

123

Response of Lm to Glycerol

Water KCl

Glycerol

Lm (H)

0.0118
0.0114
0.011
0.0106
0.0102

Time (s)

0   5000   10000   15000   20000   25000

Response of Cm to Glycerol

Glycerol

Cm (F)

2.95E-14
2.90E-14
2.85E-14
2.80E-14
2.75E-14
2.70E-14

Water KCl

Time (s)

0   5000   10000   15000   20000   25000
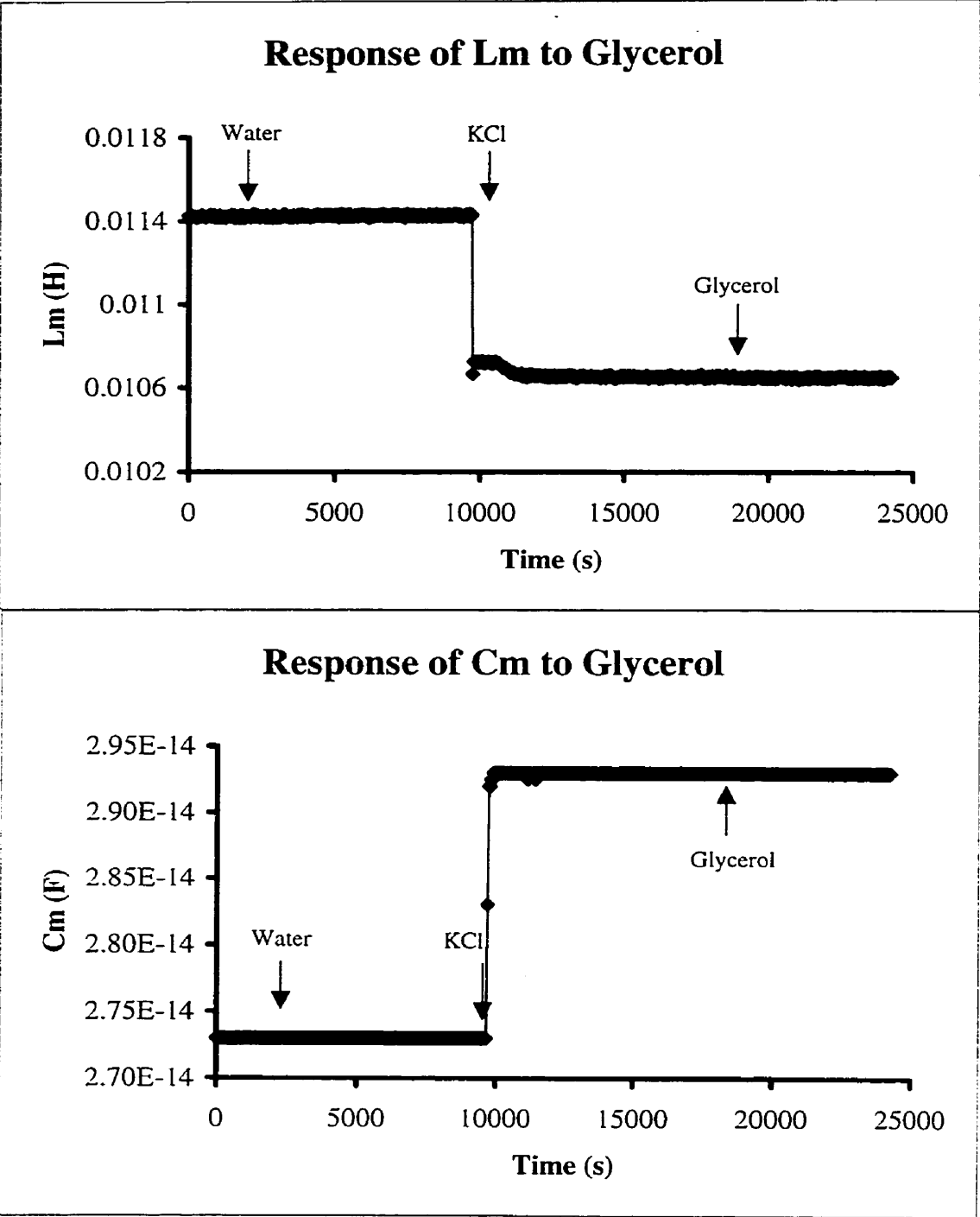
Figure 36: Responses of L_m and C_m to glycerol (0.04 M KCl solution).

124

is convenience since the NeuralWare package possesses a much more user-friendly interface. In addition, the ANNAT software is a FORTRAN-based program, located on the departmental mainframe, making it relatively inaccessible and non-portable.

Network analysis data from doped KCl solutions, of the same concentrations as those for previous studies, were subjected to the original 4,9,1 network trained on unadulterated data. The predictive results showed unacceptable accuracy in the range of 18% - 27%. This is understandable considering that two out of the four input parameters are significantly affected by the contaminant. The reduced robustness of the network is a direct consequence of the small number of input variables. The gains in speed and clarity, from the reduced complexity, do come at a price.

Weight Adjustment - A weight-adjustment scheme, analogous to that in the thiol-based study, was employed in an attempt to improve the predictive ability of the neuralnet. The weights associated with the R_m - to - hidden layer connections were reduced in magnitude incrementally until an acceptable test RMS error was achieved. Even after considerable manipulation, only a slight improvement in predictive ability was obtained, yielding errors in the 14% - 21% range. It is obvious that the poor performance of the network can be attributed to the uncorrected, glycerol induced changes in the series resonant frequency.

A logical subsequent step would be the concurrent adjustment of the F_S - related weights, in effect a two-point weight-adjustment procedure. This was executed and an improvement in the performance of the network was observed (table 6), with the predictive errors lowered to between 6% - 13%. Although this level of accuracy may

	Glycerol	Glycerol + Thiol
Predictive Error (Unadjusted)	18% - 27%	14% - 42%
Predictive Error (R_m - adjusted)	14% - 21%	-
Predictive Error (Two-point adjustment)	6% - 13%	8% - 16%

Table 6: Multi-point weight adjustment.

be acceptable for most applications, a reduction in the efficiency of the weight-adjustment procedure is apparent. A possible explanation may lie in the method utilised for the two-point adjustment process. The test RMS error was first optimised through the incremental reduction of the R_m - related weights. This was followed by optimisation using the F_S - related weights. The resulting weight factors may not be the optimum for the system in question. A better result may be obtained if the two sets of weight were to be optimised simultaneously, perhaps through the use of the simplex optimisation method. Such a process will be time consuming but the reward may be substantial, as the lowest achievable predictive errors seem to be in the 4% - 8% (table 5). Another possible explanation for the reduced efficiency may be that a two-point adjustment on a system with only four inputs could be considered to be a case of overmanipulation. This would be confirmed if simplex optimisation fails to yield an improvement. The inclusion of more input variables also may help to confirm, or disprove, this hypothesis.

Neural Analysis for a Dual-Adulterant System

Neural computation was subsequently applied to KCl solutions, adulterated by both thiol and glycerol in an attempt to elucidate and validate the two-point adjustment procedure. Once again, KCl solutions of the same concentrations as those in the earlier validation sets were doped and the resulting network analysis parameters were used to predict the concentrations. The effects of the two contaminants on the responses of the sensor are illustrated in figures 37 and 38. As expected, the combination of interferents induced significant perturbations in F_S and R_m and much

127
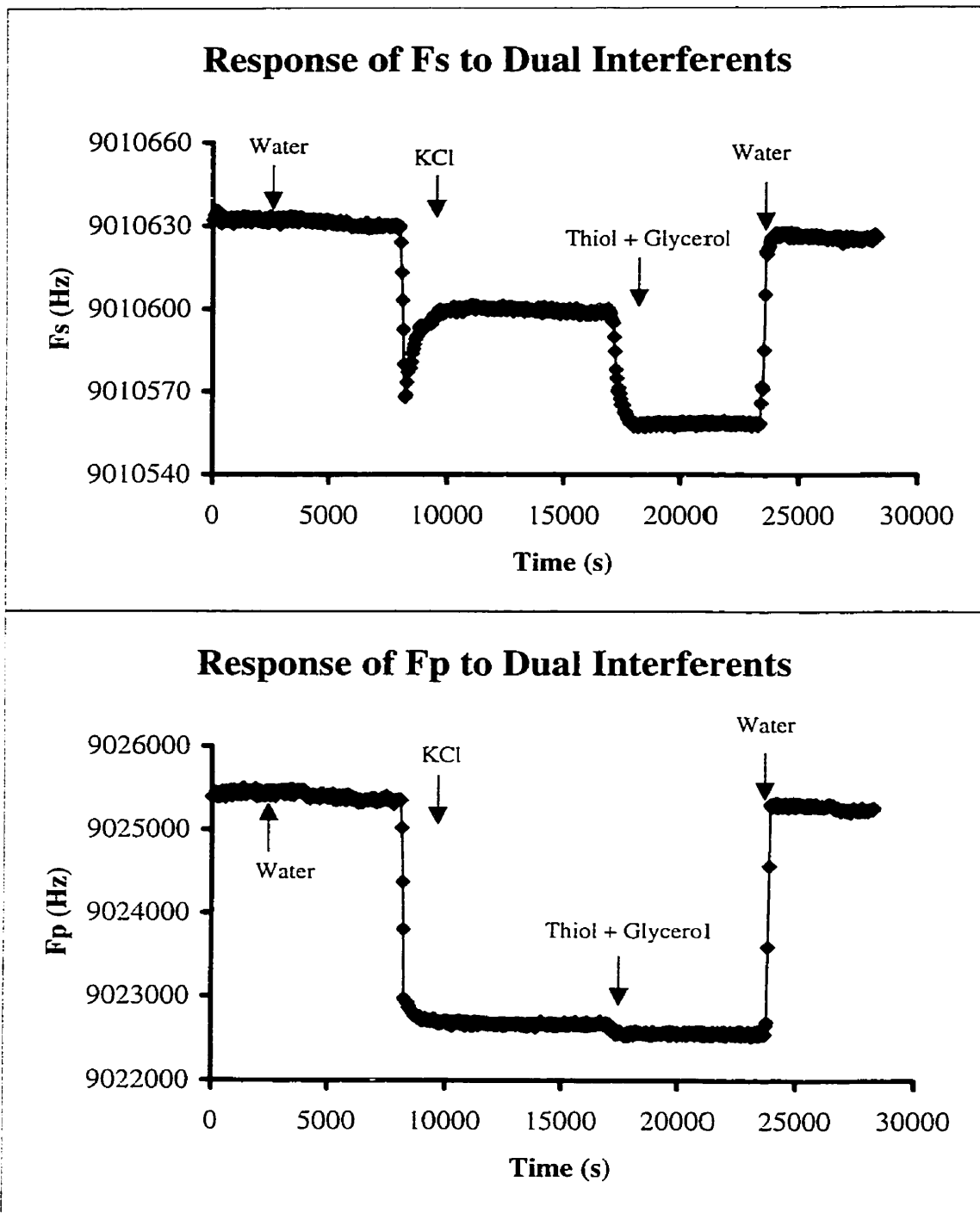
Figure 37: Responses of F_S and F_P to dual interferents.

128

Response of Co to Dual Interferents

Thiol + Glycerol

KCl

Water

Water

1.2E-11
1.1E-11
1E-11
9E-12
8E-12
7E-12

Co (F)

0   6000   12000   18000   24000   30000

Time (s)

Response of Rm to Dual Interferents

Water

KCl

Water

Thiol + Glycerol

265
259
253
247
241
235

Rm (ohm)

0   6000   12000   18000   24000   30000

Time (s)
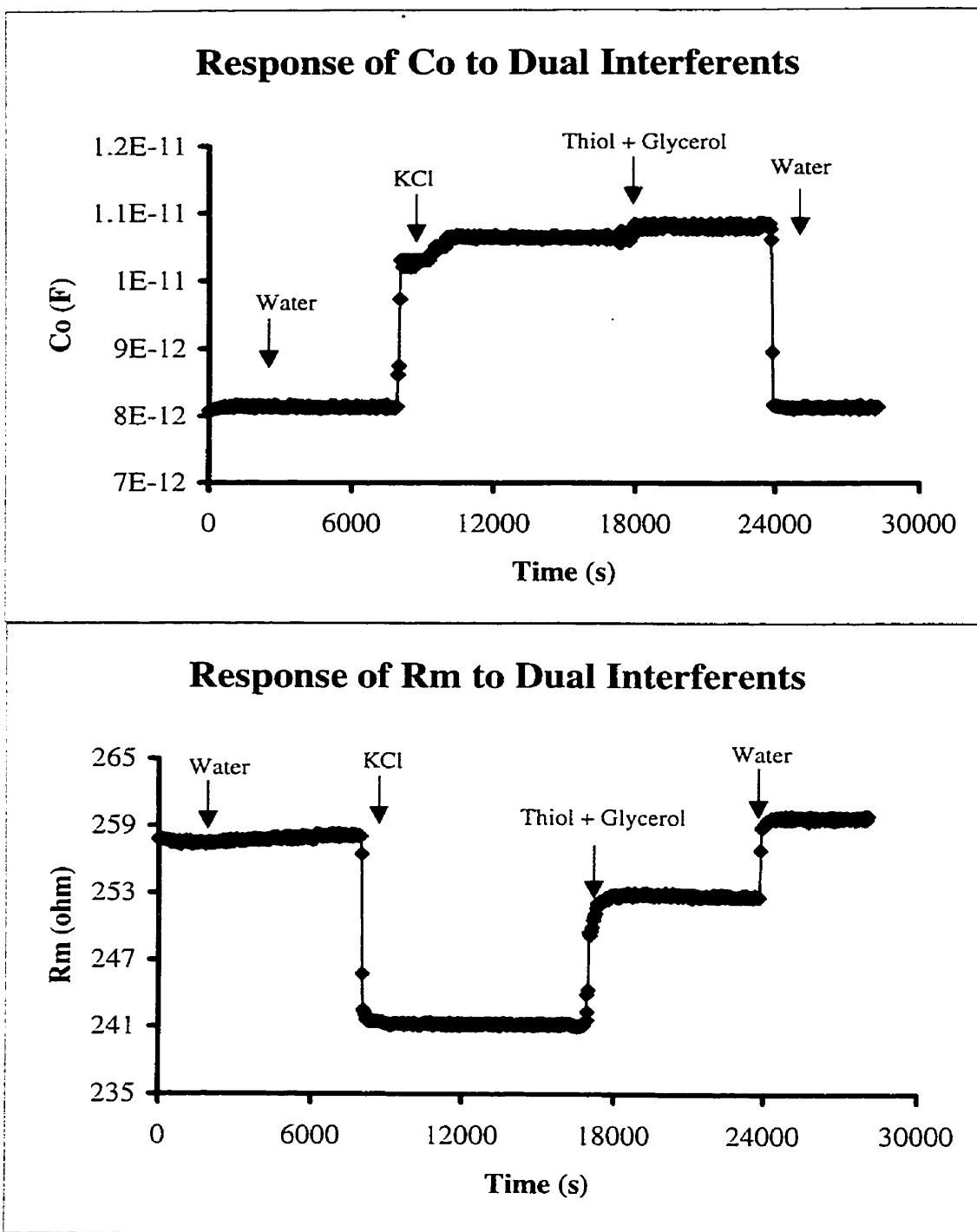
Figure 38: Responses of C_0 and R_m to dual interferents.

129

less effects on F_P and C_0. Not surprisingly, the predictive results, as derived from the unadjusted network, were less than acceptable (table 6). The combined effort of the two dopants propelled the % error above the 40% mark, rendering the neuralnet worthless. A subsequent two-point calibration procedure drastically improved the predictive accuracy of the network but the level achieved was nowhere near those of earlier adjusted networks. The reasons for this may be two-fold: inefficient optimisation of the concerned weights and / or insufficient network resources to allow for the two-point adjustment process. Once again, the application of a well-regarded optimisation routine (simplex) may produce the answer. Failing this, further investigation using larger networks may be required before the multi-point procedure can be validated.

.

DEVELOPMENT OF SENSOR PLATFORM

Surface Modification

Although the first part of this project was based on an uncoated TSM device, a chemically-selective platform is a requirement for the next step, the application of neural analysis and the accompanying calibration scheme to a real world, chemical sensing system. The logical surface-activation strategy for the gold-electroded TSM device was the self-assembly of bifunctional thiols, with carboxylic acid (-COOH) as the distal functional group. The formation of stable monolayers of alkylthiol onto a gold surface was first investigated by Nuzzo and Allara.[88] The stability was conferred by two factors: the strong affinity of the sulphur atom for zero-valent gold surfaces [89] and the orientation and stabilisation effects of Van der Waals forces between methylene groups.[90] It was demonstrated that the effects of the Van der Waals forces are most significant for alkylthiols having chain lengths of ten carbon atoms or higher.[90] Strong crystalline formation were obtained for the longer chain lengths while less ordered structures resulted from a reduction in chain length. The bonding stability of the thiol layer was demonstrated via unsuccessful solvent rinses, failed displacement utilising moieties with known strong affinities for gold (Γ, CN^-) and thermal stability up to $180°$ C.[91]

While alkylthiols will adsorb strongly to any gold surfaces, an ideal, closely packed monolayer can only be grown on a mono-crystalline surface. Since this is not the case for the TSM device, a strategy was devised to maximise the degree of functionalisation and reactivity of the monolayer. Three different surfaces were prepared: a pure monolayer of a single chain length (C_{16}), a monolayer of mixed chain lengths (C_6, C_8, C_{11} and C_{16}) and a single chain length (C_{16}) mixed with a shorter alkyl thiol spacer (C_{11}). The different self-assembly schemes are summarised in

131
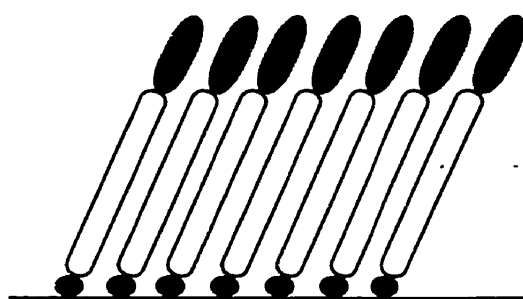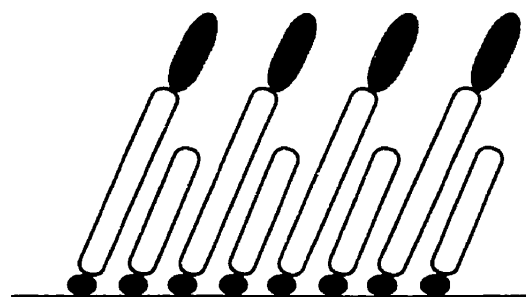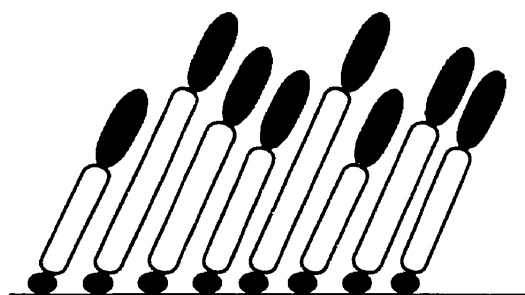
Figure 39: Thiol compositions (SLT=single-length bifunctional thiol, AT=bi-

functional and alkylthiol mixture, MT=mixed-length bifunctional thiols) .

132

figure 39. The probe molecule pentafluorophenyl hydrazine (PFPH) was employed to assess the reactivities of the respective surfaces, as the amino functional groups will form amide bonds with the distal carboxylic moieties. The probe molecule was also chosen for its fluoride content, a highly visible tag for the subsequent X-ray photoelectron spectroscopic (XPS) analysis. The high visibility of fluorine arises from the large sensitivity factor for the F1s orbital.[92] Due to the lowered reactivity of a surface-bound functional group, activation via carbodiimide chemistry was effected using dicyclohexyl carbodiimide (DCC). A catalyst, dimethyl amino pyridine (DMAP), was also utilised in the activation process to further the yield. The complete linkage reaction is presented in figure 40. It should be noted that although the mechanism of carbodiimide activation is well-understood, the role of the catalyst DMAP is still not well-defined. It is thought to catalyse the reaction as depicted or in a concerted manner, involving a transient, tri-molecular intermediate.

Table 7 depicts the results of the surface activation procedures. The blank gold surface displays a degree of contamination from air-borne carbonaceous materials. This type of contamination is present in all samples that have been exposed to ambient atmosphere, however brief the exposure may be. Since XPS is a highly surface-sensitive technique, the uppermost layer will always receive a disproportionally high percentage composition. As such, the large percentage of C1s does not reflect the true extent of contamination, which was probably quite small. In any case, this amount of carbon should be taken into account in the interpretation of any subsequent surface.

The degrees of functionalisation for the three different thiol compositions can be determined from the XPS data in table 7 (spectra can be found in appendix). From the amounts of O1s and S2p present, the pure monolayer of a single chain length (S.L.T) and the mixture of several chain lengths (M.T.) seem to have comparable,

133

carboxylated surface

(DCC)

(DMAP)

(PFPH tag)

tagged surface
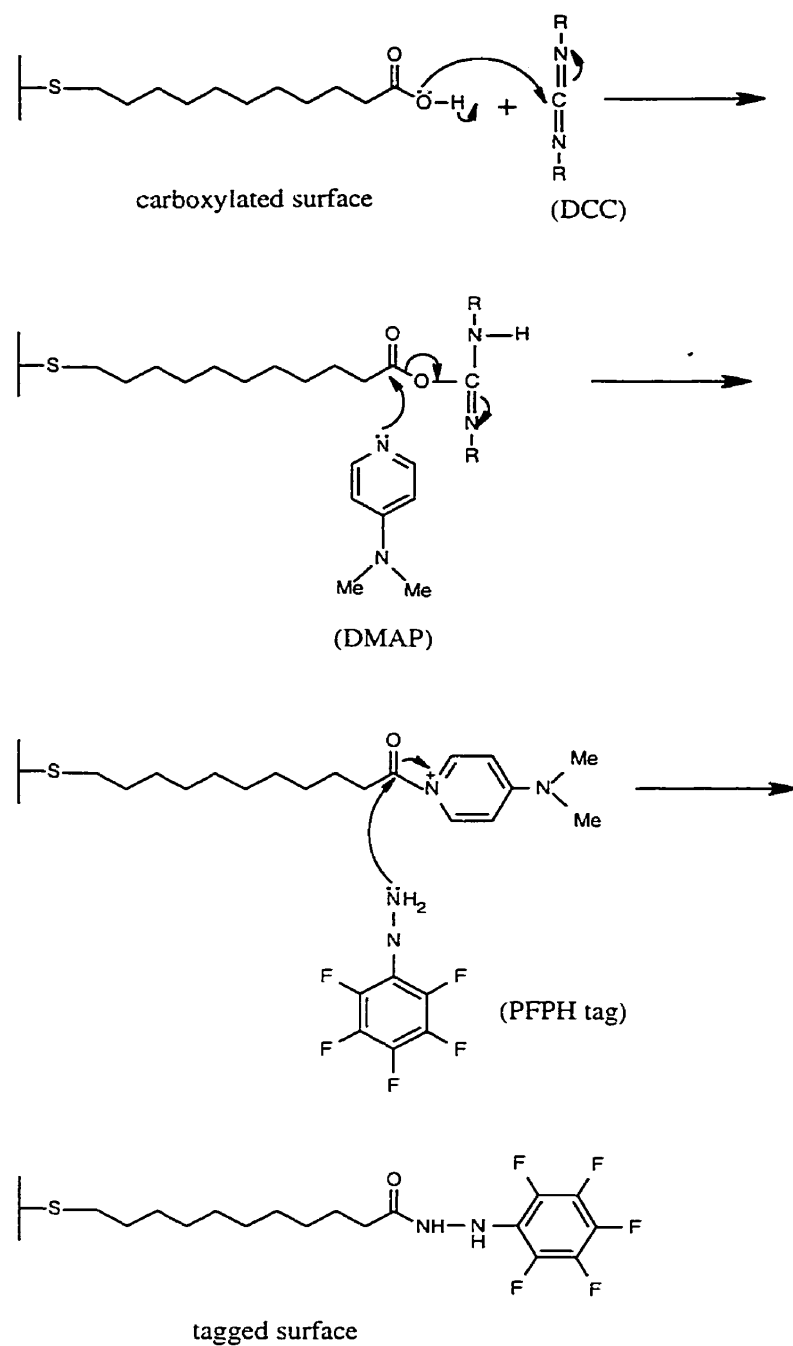
Figure 40: Schematic of carbodiimide immobilisation.

134

Surface	Fluorine (687.8 eV)	Oxygen (532.2 eV)	Nitrogen (400.5 eV)	Carbon (285.4 eV)	Sulphur (162.7 eV)	Gold (86.1 eV)
- Blank				12.5		87.5
- S.L.T.		10.5		54.3	3.5	31.9
- A.T.		7.2		64.0	1.9	26.9
- M.T.		10.7		46.4	2.5	40.5
- S.L.T. + PFPH	8.4	11.9	6.2	68.1		5.4
- A.T. + PFPH	7.5	10.0	5.5	69.9		7.0
- M.T. + PFPH	14.5	9.1	9.0	61.6		5.8
- BLANK + PFPH	0.9	4.2	0.5	27.9		66.5

- S.L.T = $HSC_{15}COOH$, A.T. = $HSC_{11}+HSC_{15}COOH$, M.T. = HSC_nCOOH (n = 5, 7, 10, 15)

- Empirically derived Sensitivity Factors were used to estimate the relative % data listed above.

- F1s = 1, O1s = 0.78, N1s = 0.54, C1s = 0.34, S2p = 0.84, Au4f = 10.8.

Table 7: XPS data for surface activation.

high loading, followed by the system of a single chain length mixed with a shorter alkyl thiol spacer (A.T.). However, a larger amount of carboxylic functional group may not necessarily translate into a higher reactivity. It has been demonstrated that a mixed monolayer, consisting of a bifunctional thiol and a shorter alkylthiol, possesses a higher reactivity, with respect to a subsequent immobilisation of the protein streptavidin, than a pure monolayer of bifunctional thiols.[93] The close-packed nature of the monolayer combined with the large size of the protein were deemed to have sterically hindered the immobilisation process, thus reducing the efficiency. The addition of the alkylthiol spacers greatly reduced the steric hindrance and the diluted bifunctional moieties were able to act as anchors for the protein.

The relationship between thiol composition and reactivity is elucidated by the reaction with the probe molecule. From the fluorine and nitrogen signals, it is evident that the surface comprising of a mixture of bifunctional thiols (M.T.) possesses the highest reactivity. This is not surprising in lieu of the fact that this surface has a high COOH loading as well as a reduced steric hindrance from the variety in chain length. It should also be noted that the steric hindrance in the immobilisation of the streptavidin is caused in part by the relatively huge size of the protein. The small size of the probe would have reduced this effect and perhaps may have even allowed it to "wriggle" into the crevices caused by the variety in chain length, thus increasing the degree of reaction. The other two surfaces have somewhat comparable reactivity, with the pure monolayer (S.L.T.) being slightly higher. This is contradictory to the streptavidin result, but may have proven the irrelevance of steric hindrance for reactions involving small molecules. A control experiment was performed, from which it can be seen that a near negligible amount of non-specific adsorption do occur in the absence of the carboxylic end group. From these results, it may be assumed that the self-assembly of a thiol mixture of different chain lengths may be the optimal

surface activation scheme for the TSM device, with respect to subsequent immobilisation of small molecules.

DNA Immobilisation and Network Analysis

The next step in the investigation involved the immobilisation of the amino-terminated, 25-mer DNA single strand (F1) to the carboxylated surface, using the protocol developed for the probe molecule (figure 40). This reaction is expected to proceed in a similar fashion, as the terminal amino group is also a primary amine of comparable nucleophilic strength. The DNA loading efficiency for all three thiol compositions can be deduced from the N1s and P2p signals in table 8. Once again, the surface derived from a mixture of bifunctional thiols (M.T.) seems to be the most reactive, followed closely by the pure monolayer (S.L.T.) and then the mixture with the alkylthiol (A.T.). The severe steric effect seen for the streptavidin is also not evident in this scenario, with the reason being the size of the incoming molecule. Although the DNA single strand is larger than the probe molecule, it is still relatively small in comparison to streptavidin. In addition, the terminal amino group is tethered to the larger DNA main body by a C_3 chain, further reducing the steric hindrance. The slight advantage of the mixture (M.T.) over the pure monolayer (S.L.T.) seems to imply a small degree of steric hindrance but it is clear that the absolute loading of reactive functional group is the determining factor.

The viability of the immobilised DNA single strand as a specific receptor was assessed through a series of hybridisation experiments. The network analysis results from these experiments are depicted in figures 41 to 44. Two curious trends can be seen from the responses of the two resonant frequencies to the hybridisation event with the complementary F2 DNA strand (figure 41). For F_s, a large drop is observed

Surface	O 1s (532.2 eV)	N 1s (400.5 eV)	C 1s (285.4 eV)	P 2p (132.1 eV)	Au 4f (86.1 eV)
- Blank			12.5		87.5
- S.L.T. + DNA	19.8	13.6	62.4	2.6	1.6
- A.T. + DNA	7.2	11.1	66.6	2.0	13.0
- M.T. + DNA	16.2	13.8	59.4	3.4	7.2
- Blank + DNA	5.6	1.5	30.4		59.2

- S.L.T = $HSC_{15}COOH$, A.T. = HSC_{11}+$HSC_{15}COOH$, M.T. = HSC_nCOOH (n = 5, 7, 10, 15)

- Empirically derived Sensitivity Factors were used to estimate the relative % data listed above.

- F1s = 1, O1s = 0.78, N1s = 0.54, C1s = 0.34, S2p = 0.84, Au4f = 10.8.

Table 8: XPS data for DNA immobilisation.

with the injection of the complementary F2 strand, followed by a stabilisation process. When the pump was stopped to allow for completeness of hybridisation, a gradual increase is seen, which then stabilises. This seems to indicate the removal of non-specifically adsorbed F2 strands. As the flow was switched back to pure buffer, the frequency increased slightly and quickly flatlined, indicative of further removal of non-hybridised fragments. The net change in frequency is on the order of 20 Hz, comparable to that obtained for a similar hybridisation event anchored by an avidin - biotin interaction.[94] In both cases, the magnitude of change does not agree with the classical Sauerbrey mass loading effect regularly invoked for the TSM device. The greater sensitivity may be attributable to changes in density and viscosity at the solid - liquid interface.[95]

The response of the parallel resonant frequency is markedly different from that of F_S. Upon addition of the F2 component, only a slight response is observed which quickly stabilises. The system remained relatively invariant throughout the equilibration process but underwent a large decrease with the switch back to pure buffer. As explained earlier, F_P is highly sensitive to perturbations in the acoustoelectric properties at the solid - liquid interface. The switch to pure buffer may have triggered a change in the conformation or orientation of the coupled DNA. This, in turn, may have induced a change in the size or nature of the charge density at the surface-adjacent layer. The decrease in F_P is consistent with an increase in acoustoelectric coupling with the bulk solution. A thinner and/or less dense double layer would bring about such a deviation.

Examination of the other electrically sensitive parameters, C_0 and L_m, revealed identical trends (figure 42). Both variables remained unperturbed throughout the whole hybridisation event only to experience large shifts once the flow was switched back to pure buffer. The shifts for C_0 and L_m also appear to reflect either an increase

Response of Fs to Hybridisation

Response of Fp to Hybridisation
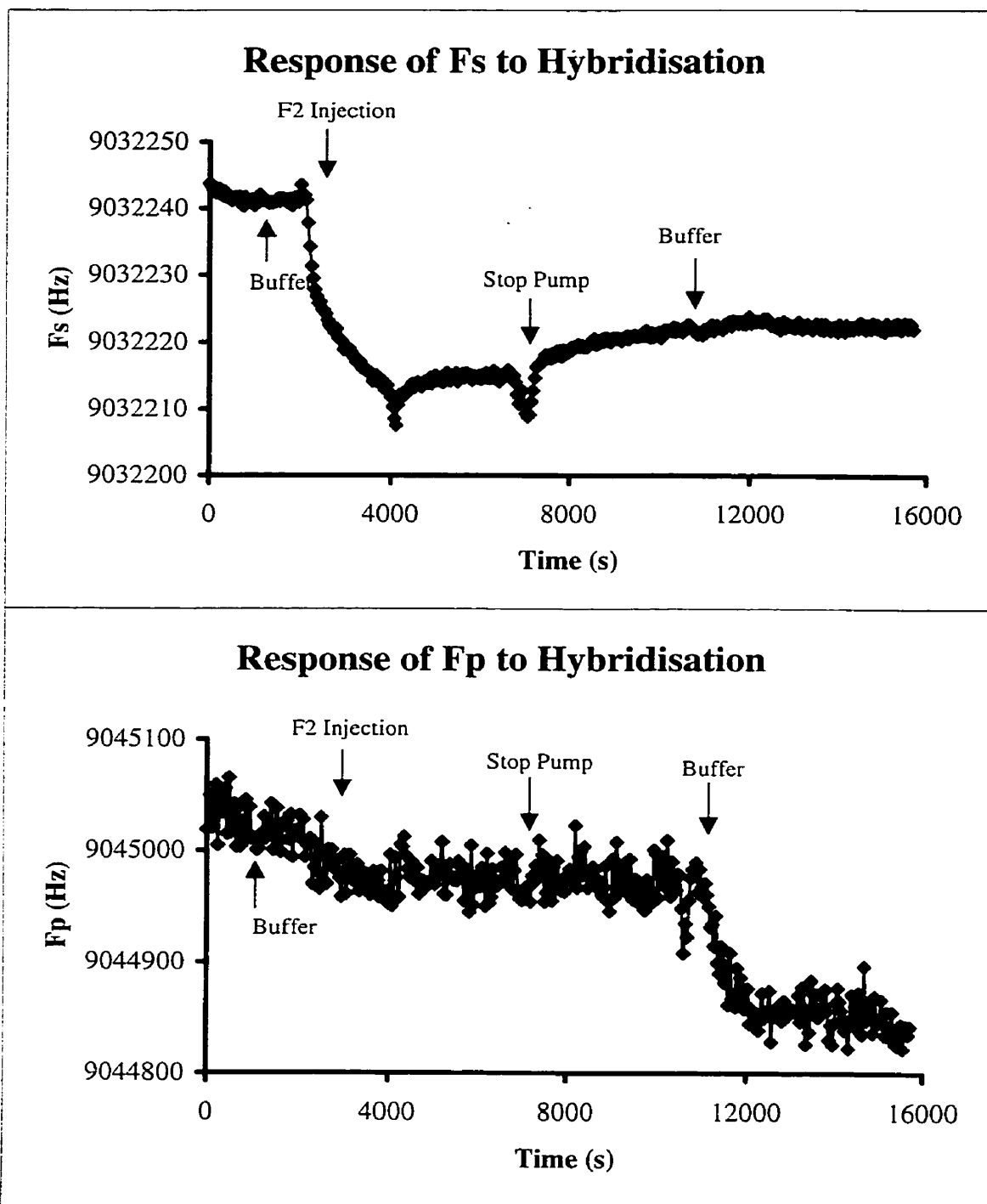
Figure 41: Responses of F_S and F_P to hybridisation (DNA concentrations of 167 µM).

140

Response of Co to Hybridisation

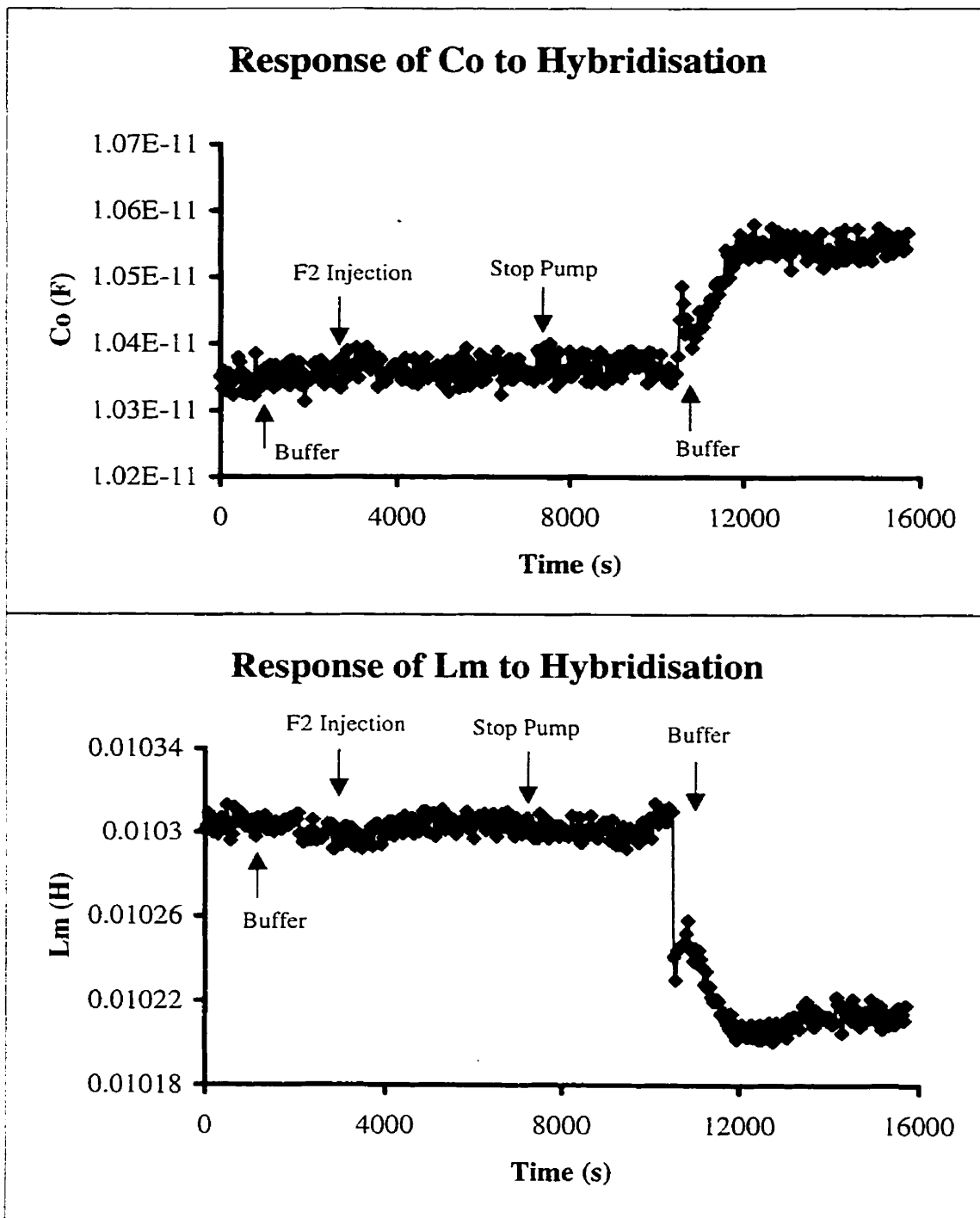
Response of Lm to Hybridisation

Figure 42: Responses of C_0 and L_m to hybridisation (DNA conc. of 167 μM).

141

in acoustoelectric coupling or a decrease in the shielding experienced by the TSM sensor. On the other hand, the trends exhibited by parameters influenced by mass and viscous effects, R_m and Q-value, follow that of the series resonant frequency (figure 43). In both cases, the largest change is induced by the hybridisation event itself and not by a secondary conformational change. The changes brought about by the equilibration and washing processes mirror that of F_S, gradual and relatively minor. A control experiment was performed in which the target DNA was exposed to a non-complementary fragment, F0 (figure 44). An initial drop in F_S is seen upon exposure to the F0 solution but once the pump was stopped, the frequency quickly rose to a level just below that of the initial buffer. The washing procedure restored the frequency to pre-hybridisation level. The transient initial drop may be attributed to the non-specific interaction between the non-complementary strands while the slight decrease prior to washing can be ascribed to non-specific adsorption of the DNA probes.

The viability of a DNA receptor system, immobilised via bifunctional thiol linkers, has been demonstrated. The sensitivity obtained is comparable to that of a system anchored by the well established avidin - biotin coupling. This is indicative of a similar packing density or an equal amount of reactive sites. Furthermore, network analysis has been shown to be perceptive to both the actual hybridisation event and an apparent secondary capacitive-based relaxation process. The multi-parametric nature of network analysis allows for the simultaneous monitoring of the acoustoelectric coupling as well as the mass and viscous coupling effects.
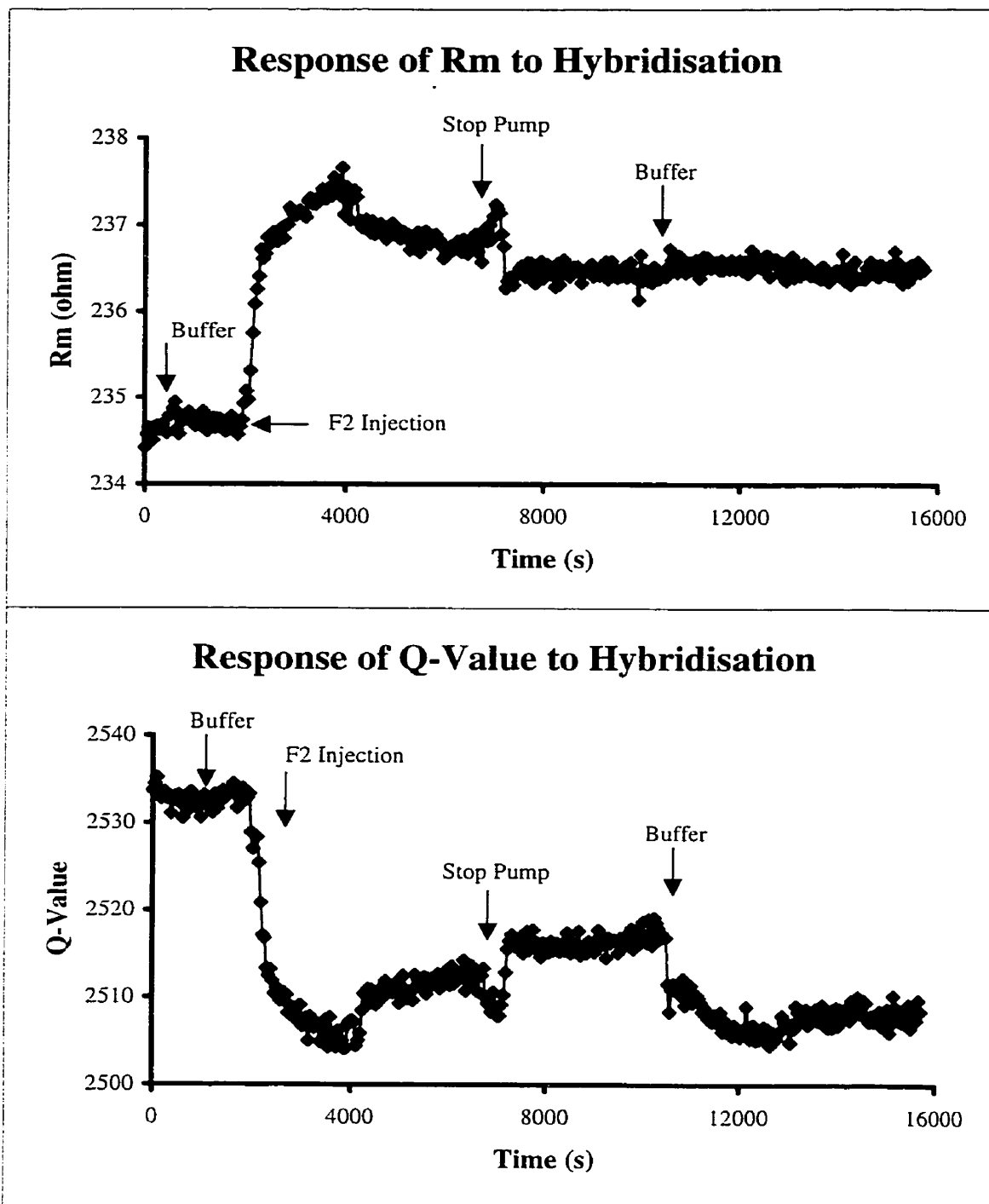
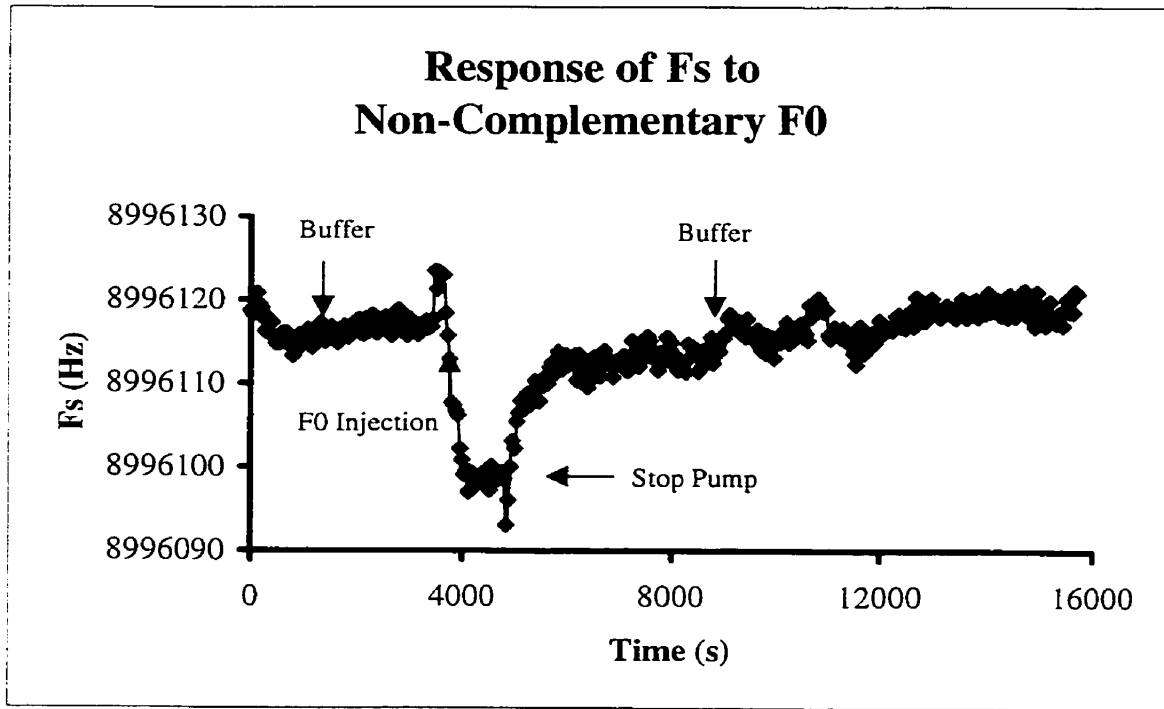Figure 43: Responses of R_m and Q-value to hybridisation.

Figure 44: Response of F_S to non-complementary F0.

Atomic Force Microscopy Analyses

Atomic force microscopy (AFM) was employed to further characterise the various surfaces produced in the immobilisation of the surface-bound DNA receptors. The tapping mode of operation was chosen to minimise the lateral forces acting on the surface, as the contact time between the probe and the sample is reduced relative to the contact mode of operation.[96] As the tip is lightly oscillated vertically near its mechanical resonance frequency, the amplitude of oscillation is reduced via contact with features on the surface. The height of the sample platform is adjusted to maintain a constant amplitude of probe oscillation. This is the "height" mode of operation, which gives a topographical relief of the surface. The brighter the colour in the image, the larger the vertical displacement of the feature.

Images were simultaneously acquired in the phase mode, in which the motion of the probe is characterised by its phase relative to the driving oscillator.[97] The phase lag between the excitation signal and the probe response changes when the tip is in contact with chemical heterogeneities. Thus, phase imaging is sensitive to the differences in phase between domains on the surface. Phase shifts are registered as bright and dark regions in a phase image. However, twisting motions of the probe, as it encounters steep features also give rise to a phase shift. Thus, it is possible to observe contrast in a phase image even in the absence of chemical heterogeneity on the surface. However, phase imaging is relatively insensitive to surface roughness, which may hinder domain identification in height imaging. The two techniques are thus complimentary and provide the maximum characterisation of a surface.

AFM results for the control TSM surface and the thiolated surface are illustrated in figure 45a and 45b, respectively. The left image of each figure portrays the height image while the right one depicts the phase image. The height image of the control

(a)

(b)

0   500 nm 0   500 nm

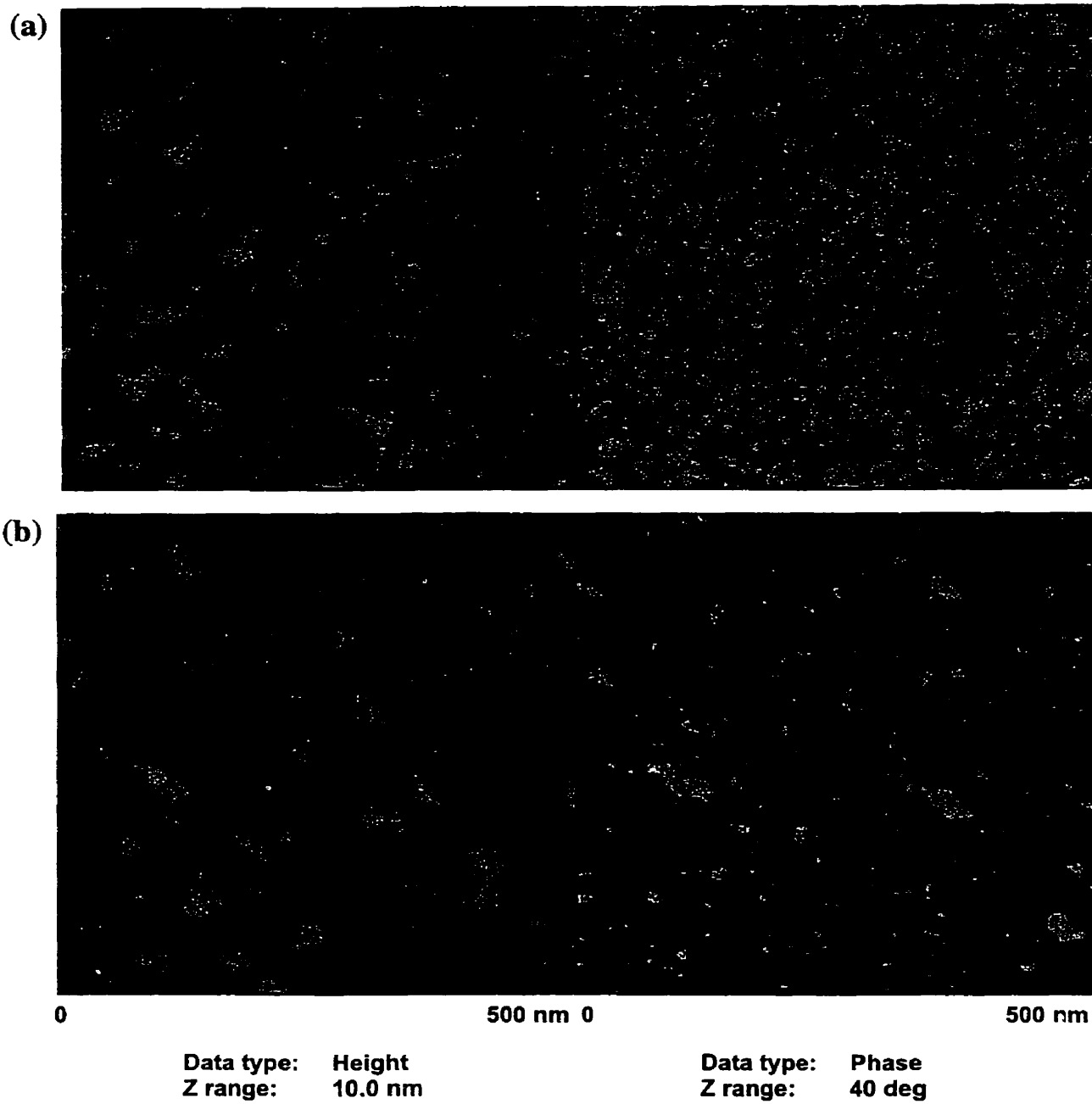| Data type: | Height | | Data type: | Phase |
| Z range: | 10.0 nm | | Z range: | 40 deg |

Figure 45: (a) Height and phase images of the control TSM surface.
(b) Height and phase images of the thiolated TSM surface.

146

surface describes a relatively rough topography with numerous large, irregular clusters, of medium brightness (height), which may represent gold islands. A few extra bright (high) features seem to be artefacts, as suggested by the regularities in their shape and size. The corresponding phase image seems to reinforce this interpretation. The contrast in phase between the gold clusters and the possible artefacts is more evident than in the topographic image. The highest (brightest) objects were clearly shown to be circular in shape, supporting the theory that they are physisorbed specks of contamination. The structures thought to be gold clusters were confirmed to be irregular in shape, lending credence to their supposed identity. The difference between the gold clusters and the darker areas, assumed to be the underlying gold surface, may be attributable to the topographic component of phase imaging. The possibility exists that the darker domain may be the chromium underlayer. Both images were obtained at relatively low resolutions (500 nm x 500 nm) and the smallest visible structures are on the order of 10 nm. Attempts were made to obtain molecular resolution but it soon became evident that this was not feasible for such a rough surface. However, the combination of topographic and phase imaging may yield information about subsequent immobilised overlayers.

The distributions in size and height of the features on an AFM image can be ascertained via cross-sectional analysis and surface roughness analysis, respectively. In the former method, the diameters for several features, of similar composition, are measured and averaged to arrive at a mean cross-sectional value, with an accompanying measure of variation. Figure 46 depicts the cross-sectional analysis for the control TSM surface. The analysis was performed on a 500 x 500 nm spot size, with a vertical resolution of under 1nm. Three small circular structures, all hypothesised to be physisorbed artefacts, were chosen as subjects. The mean value (L) was found to be 32 nm with a RMS of 2 nm. Using the same method, the

Section Analysis

L	32.227 пм
RMS	1.745 пм
Ic	DC
Ra(lc)	1.541 пм
Hмэк	4.951 пм
Rz	4.951 пм
Rz Cnf	2
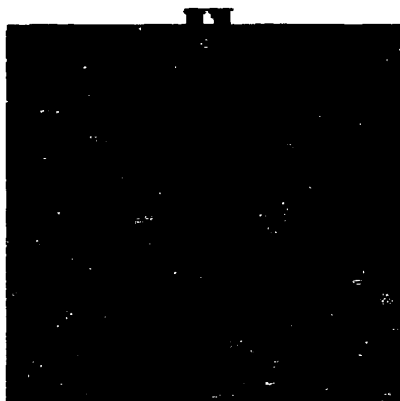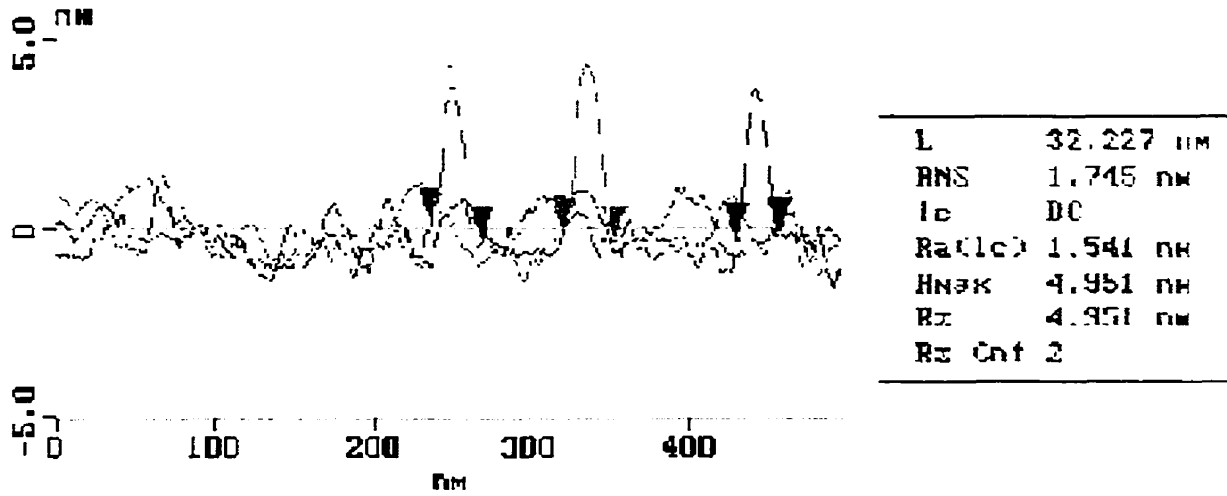
Surface distance	34.809 пм
Horiz distance(L)	32.209 пм
Uert distance	0.479 пм
Angle	0.827 deg
Surface distance	34.082 пм
Horiz distance	32.227 пм
Uert distance	0.195 пм
Angle	0.347 deg
Surface distance	29.082 пм
Horiz distance	27.341 пм
Uert distance	0.111 пм
Angle	0.233 deg

Figure 46: Cross-sectional analysis of the control TSM surface.

148

Roughness Analysis

Image Statistics

Img. Z range	104.06	nm
Img. Mean	-0.003	nm
Img. Raw mean	-34.104	nm
Img. Rms (Rq)	2.450	nm
Img. Ra	1.361	nm

Box Statistics

Z range
Mean
Raw mean
Rms (Rq)
Mean roughness (Ra)
Box x dimension
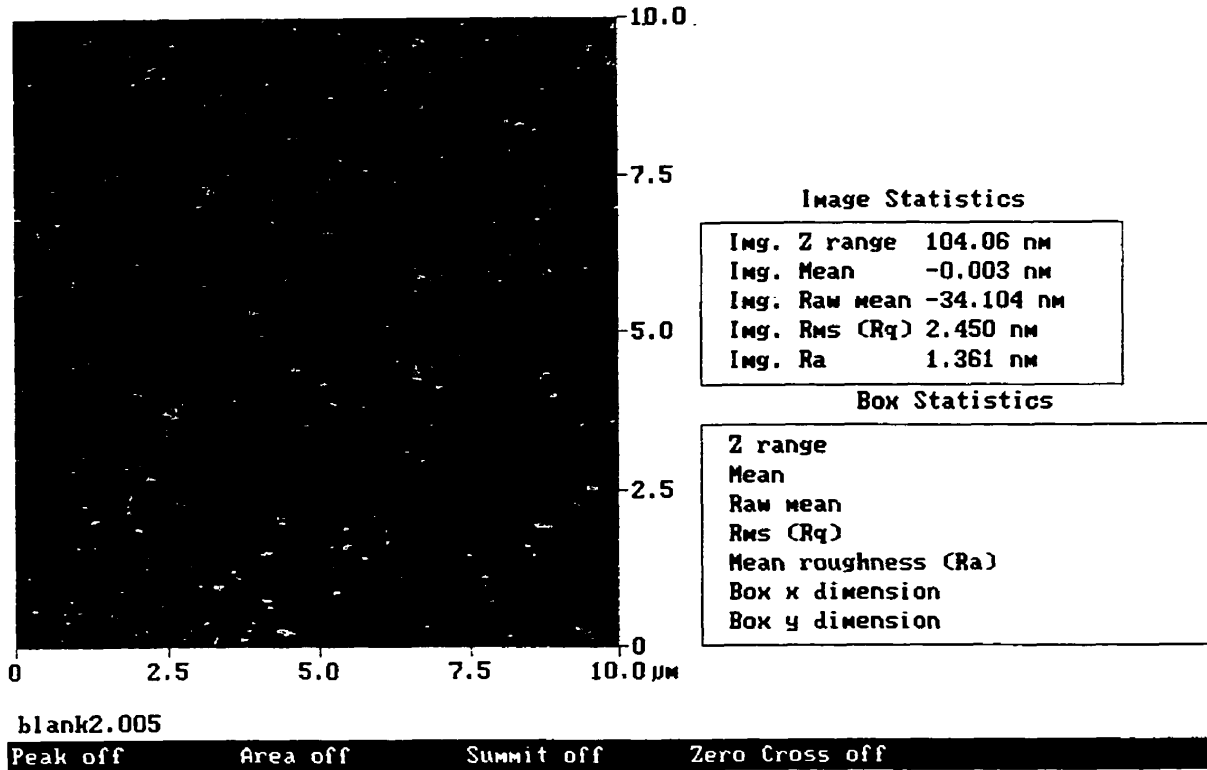Box y dimension

blank2.005

Figure 47: Roughness analysis of the control TSM surface.

149

diameters of the gold clusters were found to be in the range of 30 to 50 nm.

An analogous procedure is performed on the vertical scale of an AFM image to render a measure of surface roughness. The vertical displacements of the five highest peaks and the five lowest valleys, relative to a mean plane, are analysed in a similar fashion to yield an average deviation from the mean plane and an accompanying RMS value. The roughness analysis of the control surface, illustrated in figure 47, was performed on a 10 μm x 10 μm area. The large scale image clearly shows the bright white (highest) specks to be adsorbed contaminants. Numerous deep (dark) fissures are seen, crisscrossing the surface. One large and several smaller defect sites (dark spots) are also observed. The range of vertical displacement (Z) was found to be 104 nm, giving an average deviation (Ra) of 1.4 nm and a RMS of 2.4 nm. These values agree well with the polish specification provided by the manufacturer of the crystal (< 1 μm) but the large range value suggests relatively large deviations among the peaks and valleys on the surface *i.e.* a microscopically uneven surface. Once again, this corresponds well with the contrast in height observed in the height image.

Figure 45b depicts the height and phase images of the thiolated TSM surface (pure monolayer of bifunctional C_{16} thiol). A smearing effect is seen for the topographic image to the left. The large gold clusters and the circular specks can still be seen but there seems to be fewer of them and the differences in vertical displacements are less distinct. This may correspond with a layer of thiol being immobilised on the surface, masking the structures to a degree. A profound difference can be seen in the phase image, relative to the control surface. The surface is nearly homogenous in phase, with the artefacts clearly shown to be adsorbed on top. Once again, this is consistent with the presence of a dense layer of thiol on the gold surface. The white horizontal streaks are thought to be artefacts of the movement of the probe tip as it scans across the sample. This is often observed when the tip is contaminated with relatively labile

(a)

(b)

0   500 nm 0   500 nm

Data type: Height   **Data type:** Phase
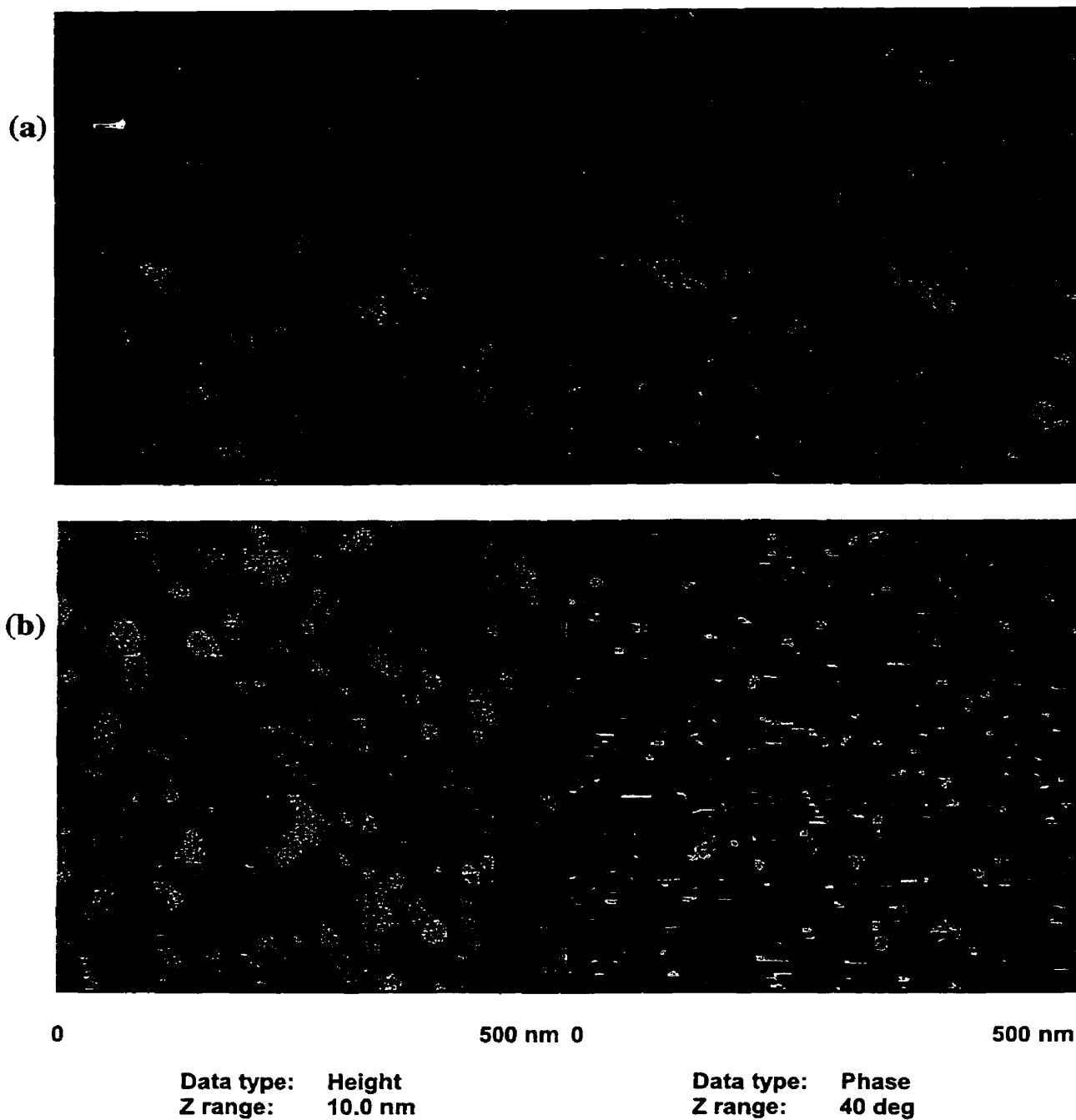**Z range:**   10.0 nm   **Z range:**   40 deg

Figure 48: (a) Height and phase images of the thiolated TSM surface.
(b) Height and phase images of the DNA-immobilised
TSM surface.

surface-bound entities. Cross-sectional analysis revealed that the size of the circular artefacts remained unchanged (29 nm ± 1 nm) while the gold clusters were shown to be larger, ranging from 40 to 80 nm. This is further evidence of the immobilisation of a layer of thiol on the gold clusters. Roughness analysis results indicated a range (Z) of 18 nm, an average deviation of 1 nm and a RMS of 1.4 nm. All of these values are significantly less than those for the control surface, further proof of a masking action by the thiol layer.

The AFM images for the thiolated surface are juxtaposed with the height and phase images of the DNA-immobilised surface for ease of comparison (figure 48). In the topographic image, the smearing effect seen for the thiolated surface has been reversed. The features are quite distinct and there are larger variations in height among them. This may be evidence of an incomplete coverage of DNA molecules on top of a thiol layer. The phase image seems to support this depiction. A homogenous underlayer is clearly seen, with numerous outcrops of a material of difference phase. The bright white circular contaminants are also seen, along with the streaky tip-induced artefacts. Cross-sectional analysis pointed to a reduction in the size of the features to the range of 30 to 60 nm while roughness measurements yielded a range of 99 nm, a mean deviation of 1.8 nm and a RMS of 3.6 nm. The increase in roughness corresponds well to an incomplete monolayer of DNA molecules.

AFM imaging has proven to be very useful in the characterisation of the immobilisation process. All though molecular resolution was not possible with the relatively rough TSM surface, height and phase imaging combined to give important supporting evidence at each step of the immobilisation procedure. The employment of an atomic flat surface, such as cleaved mica, as the substrate will help to improve the resolution and provide further information on the thiol-DNA linkage system.

CONCLUSION

In summary, artificial neural networks were applied successfully to the modelling of the KCl system. Excellent predictive ability, over a concentration range of three orders of magnitude, was obtained for a network with as few as four input parameters. The neuralnet was subjected to KCl samples that had been deliberately contaminated with a short-chain thiol, designed to affect only one of the input variable. The predictive ability of the network suffered a decline but the accuracy remained at an acceptable level. A weight adjustment scheme was developed to calibrate the network for the presence of the interferent. The magnitudes of the weights, which are related to the affected input, were adjusted downward. Essentially, the contributions from the affected nodes toward the output of the network were reduced. The predictive errors achieved by the adjusted network were found to be comparable to those for the unadulterated samples. In effect, the adjustment process had lessened the deleterious effect of the interferent in a manner roughly analogous to the calibration process for a pH meter.

The effect of a different interferent on the KCl system was investigated. The contaminant, glycerol, was designed to affect not one, but two input parameters. The predictive ability of the unadjusted network was found to be unacceptable. Weight-adjustment on one of the affect inputs did not enhance the performance greatly but a two-point adjustment improved the predictive accuracy to an acceptable level. However, the network was not totally calibrated for the effect of glycerol by the adjustment process. This can be attributed to an inefficient weight adjustment process and/or an over-extension of the network's resources. The inclusion of more relevant parameters, other than the chosen four, may help to lower the errors and improve

robustness, but at greater computing costs. An investigation was also performed on a dual-adulterants system, using both thiol and glycerol. Two of the four input parameters were affected and the results followed the same trend as those for just glycerol, but with even larger magnitudes of error. The degradation in performance can be attributed to the combined effect of the two contaminants.

A parallel development of a chemically selective sensor platform was effected in preparation for the application of neural analysis to a real-world chemical sensing system. This involved the activation of the gold electrodes on the TSM device via the self-assembly of bifunctional thiols, followed by reactivity assessment using a probe molecule, pentafluorophenyl hydrazine. The optimal thiol composition, with respect to thiol loading and surface reactivity, was found to be a mixture of several bifunctional thiols with different chain lengths (C_6, C_8, C_{11} and C_{16}). The subsequent immobilisation of an amino-terminated, 25-mer DNA single strand was performed utilising carbodiimide chemistry, catalysed by dimethylamino pyridine. Once again, the optimal surface was confirmed to be that activated by a mixture of bifunctional thiols.

The viability of the immobilised DNA was confirmed via hybridisation experiments with complementary fragments, monitored by network analysis. The net change in F_S, upon hybridisation, was found to be comparable to that obtained with a system anchored by the avidin-biotin coupling process. Furthermore, it was discovered that the multi-parametric nature of network analysis allows for the simultaneous characterisation of the actual hybridisation event and a secondary capacitive-based conformational change. The changes in mass and viscous loading, caused by the initial hybridisation, was tracked by F_S and R_m while an apparent subsequent conformational change, reflected in the acoustoelectric properties of the

154

sensor, was monitored by the electrically-sensitive parameters, F_P, C_0 and L_m.

AFM was also employed in the characterisation of each step of the immobilisation process. The combination of height and phase imaging proved extremely useful in providing corroborating evidences pertaining to the presence of the thiol and DNA layers. The application of this procedure to a layer of gold that has been sputtered onto an atomically flat surface may improve resolution to the molecular level.

Overall, there is still much to be done in both aspects of this investigation. The weight-adjustment procedure must be elucidated further before it can be applied to a real-world sensor. This would involve the creation of a larger network, making use of six or more input parameters. The inherent resources of such a network would be greater, conferring an increase in robustness and a higher capacity towards multi-point weight manipulation. The immobilisation procedure must also be optimised for greater signal sensitivity towards the hybridisation event. Although the change in frequency obtained from the DNA coupling is close to that achieved using the avidin-biotin interaction (25 Hz vs. 40 Hz), it must be improved to be competitive. This can be accomplished via higher receptor loading or increased reactivity. With respect to receptor loading, all steps of the procedure must be characterised fully with respect to surface coverage to determine the deficiencies in the process. This can be done through angle-resolved XPS and more rigorous AFM imaging. Different thiol compositions should also be attempted since the 1:1 ratio used for the thiol mixtures may not be the optimum, with respect to reactivity. Furthermore, an increase in signal sensitivity may be achieved simply by increasing the temperature of hybridisation. It is a well-known fact that a higher temperature, with the appropriate annealing procedure, will lead to greater hybridisation efficiency. This, in turn, should result in greater signal sensitivity but care must be taken since the higher temperature may also affect the network analysis signals in a non-specific manner.

155

GLOSSARY

A	piezoelectrically active area of a TSM device.
A_1	numerical constant
A_2	numerical constant
A_3	numerical constant
A_4	numerical constant
A_5	numerical constant
A_6	numerical constant
A_7	numerical constant
A_8	numerical constant
B_1	numerical constant
C	numerical constant
C_1	numerical constant
C_o	static capacitance of the equivalent circuit
C_{66}^{-}	effective complex shear modulus of a AT-cut piezoelectric quartz crystal
f_o	fundamental series resonant frequency of a quartz crystal
f_m	frequency at maximum impedance
f_s	series resonant frequency
G	complex modulus of a polymer film
K	dielectric constant of the quartz
K^2	electromechanical coupling factor
L_1	unperturbed motional inductance
Δm_L	mass per unit area of equivalent liquid layer

156

N	harmonic number
$P\text{-}P_m$	pressure difference between two sides of quartz crystal
R_m	motional resistance
T_f	thickness of a polymer immobilized on a TSM sensor surface
Z_f	Acoustic impedance of the film
ε	piezoelectric stress constant
ε_{22}	dielectric constants for a quartz crystal
ε_L	dielectric constant of a liquid
κ_L	specific conductivity of a liquid
η_L	bulk viscosity of the liquid
μ_L	bulk density of a liquid
μ_Q	elastic constant of quartz (piezoelectrically stiffened)
ρ_f	polymer density
ρ_L	density of liquid
ρ_Q	density of quartz
σ	solution conductivity
ω	angular frequency
ω_s	angular series frequency

REFERENCES

1. Robb E. W. and Munk M.E., *Mikrochim. Acta* 1990, 131.

2. Anker L. S. and Jurs P. C., *Analytical Chemistry* 1992, **64**, 1157.

3. Lin C. W., La Manna J. C. and Takefuji, Y., *Biological Cybernetics* 1992, **67**, 303.

4. Tanabe K., Tamura T. and Uesaka H., *Applied Spectroscopy* 1992, **46**, 807.

5. Hoskins J. C. and Himmelblau D. M., *Comput. Chem. Eng.* 1988, **12**, 881.

6. Svozil D., Kvasnicka V. and Pospichal J., *Chemom. Intell. Lab. Syst.* 1997, **39**, 43.

7. Bhat N. and McAvoy T., *J. Comput. Chem. Eng.* 1990, **14**, 573.

8. Kneller D. G., Cohen F. E. and Langridge R., *J. Mol. Biol.* 1990, **214**, 171.

9. Bohr H., Bohr J., Brunak S., Cotterill R. M. J., Fredholm F., Lautrup B., Olsen O.H. and Petersen S. B., *FEBS Lett.* 1990, **261**, 43.

10. Bos M., Bos A. and Van der Linden W. E., *Anal. Chim. Acta* 1990, **233**, 31.

11. Walt D. R., Dickinson T. A., Healy B. G., Kauer J. S. and White J., *Proc. SPIE-Int. Soc. Opt. Eng.* 1995, **2508**, 111.

12. Ganesh C., Steele J. P. H., Zhang H., Mishra D. and Jones J., *Int. SAMPE Symp. Exhib.* 1994, **39**, 883.

13. Fujii S., Kamada T., Hayashi S., Tomita Y., Takayama R., Hirao T., Nakayama T. and Deguchi T., *Proc. SPIE-Int. Soc. Opt. Eng.* 1995, **2332**, 612.

14. Nanto H., Tsubakino S., Kawai T., Ikeda M., Kitagawa S. and Habara M., *J. Mater. Sci.* 1994, **29**(24), 6529.

15. Wang X., Yee S. and Carey P., *Sens. Actuators B* 1993, **13**(1-3), 458.

16. Tonoike M., *Cosmet. Toiletries* 1990, **105**(6), 49.

17. Sah W. J., Lee S. C., Tsai H. K. and Chen J. H., *Appl. Phys. Lett.* 1990, **56**(25), 2539.

18. Chang S. M., Muramatsu H. and Karube I., *Sens. Mater.* 1995, **7**(1), 13.

19. Hivert B., Hoummady M., Henrioud J. M. and Hauden D., *Sens. Actuators B* 1994, **19**(1-3), 645.

20. Moriizumi T., Nakamoto T. Sakuraba, Y., *Olfaction Taste XI, Proc. Int. Symp., 11th* 1994, 694.

21. Nanto H., Tsubakino S., Ikeda, M. Endo F., *Sens. Actuators B* 1995, **B25**(1-3), 794.

22. Pisanelli A., Qutob A. A., Travers P., Szyszko S. Persaud, K., *Life Chem. Rep.* 1994, **11**(2), 303.

23. Curie J. and Curie P., *Bull. Soc. Min. France*, 1880, **3**, 90.

24. Cady W. G., *Piezoelectricity, Vol. 1*, Dover Publications, New York, 1964, pg. 4.

25. Hankel W. G., *Abh. Sächs* 1881, **12**, 547.

26. Thompson M. and Stone D. C., *Surface-Launched Acoustic Wave Sensors, Vol. 144 in Chemical Analysis*, Ed. Winefordner J. D., Publ. John Wiley & Sons, Inc., Toronto, Canada, 1997.

27. Sauerbrey G., *Z. Phys.* 1959, **155**, 206.

28. Lide D. R., *Handbook of Chemistry and Physics*, 73rd Edition, CRC Press.

29. Miller J. G. and Bolef D. I., *J. Appl. Phys.* 1968, **39**, 4589.

30. Lu C. S. and Lewis O., *J. Appl. Phys.* 1972, **43**, 4385.

31. Lu C. S. and Czanderna A. W., *Applications of Piezoelectric Quartz Crystal Microbalances*, Elsevier, Amsterdam, 1984.

32. Nomura R. and Minemura A., *Nippon Kagaku Kaishi*, 1980, **198**, 1621.

33. Nomura T. and Okuhara M., *Anal. Chim. Acta*, 1982, **142**, 281.

34. Bruckenstein S. and Shay M., *Electrochim. Acta* 1985, **10**, 1295.

159

35. Kanazawa K. K. and Gordon J. G., *Anal. Chim. Acta* 1985, **175**, 99.

36. Reed C. E., Kanazawa K. K. and Kaufman J. H., *J. Appl. Phys.* 1990, **68**, 1993.

37. Martin S. J., Granstaff V. E. and Frye G. C., *Anal. Chem.* 1991, **63**, 2272.

38. Haeusler K. E., Grzegoezewski A., Jackel A. and Pietrucha J., *Ber. Bunsen-Ges. Phys. Chem.* 1988, **92**, 1218.

39. Hager H. E., *Chem. Eng. Commun.* 1986, **43**, 25.

40. Shana Z. A. and Josse F., *Anal. Chem.* 1994, **66**, 1955.

41. Muramatsu H., Tamiya E. and Karube I., *Anal. Chem.* 1988, **60**, 2142.

42. Yao S. Z. and Zhou T. A., *Anal. Chim. Acta* 1988, **212**, 61.

43. Schumacher R. , Gordon J. and Melroy O., *J. Electroanal. Chem.* 1987, **216**, 127.

44. Martin S. J., Frye G. C., Ricco A. and Senturia S. D., *Anal. Chem.* 1993, **65**, 2910.

45. Yang M. and Thompson M., *Langmuir* 1993, **9**, 1990.

46. Kipling A. L. and Thompson M., *Anal. Chem.* 1990, **62**, 514.

47. Rajoković Lj. V., Ćavić-Vlasak B. A., Ghaemmaghami V., Kallury K. M. R., Kipling A. L. and Thompson M., *Anal. Chem.* 1991, **63**, 615.

48. Thompson M., Dhaliwal G. K., Arthur C. L. and Calabrese G. S., *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.* 1987, UFFC-**34**, 127.

49. Yang M., Thompson M. and Duncan-Hewitt W. C., *Langmuir* 1993, **9**, 802.

50. Duncan-Hewitt W. C. and Thompson M., *Anal. Chem.* 1992, **64**, 94.

51. Krausz A. S. and Eyring H., *Deformation Kinetics*, 1973, Wiley, New York.

52. Mitchell D. J., Ninham B. W. and Palithorpe B. A., *J. Colloid Interface Sci.* 1978, **64**, 194.

53. Sullivan D. E. and Still G., *J. Chem. Phys.* 1979, **69**, 5450.

54. Croxton C. A. Ed., *Fluid Interfacial Phenomena*, Publ. J. Wiley and Sons, Chichester, U.K., 1986.

55. Derjaguin B. V. and Churaev N. V., in *Fluid Interfacial Phenomena*, Ed. Croxton C. A., Publ. J. Wiley and Sons, Chichester, U.K., 1986, pg. 663.

56. Tolstoi D. M., *Dokl. Acad. Nauk. USSR* 1952, **85**, 1089.

57. Blake T. D., *Colloids Surf.* 1990, **47**, 135.

58. Gee M. L., Mc Guiggan P. M., Israelachvili J. N. and Hornola A. M., *J. Chem. Phys.* 1990, **93**, 1895.

59. Krim J., *Bull. Am. Phys. Soc.* 1988, **33**, 436.

60. Hayward G., *Anal. Chim. Acta* 1992, **264**, 23.

61. Ferrante F., Kipling A. L. and Thompson M., *J. Appl. Phys.* 1994 , **76**, 3448.

62. Hayward G. and Thompson M., *J. Appl. Phys.* 1998, **83**, 2194.

63. Adamson A. W., Physical Chemistry of Surfaces, 4th ed., Wiley, NY, 1982.

64. Rodahl M., Fredrik H. and Kasemo B., *Anal. Chem.* 1996, **68**, 2219.

65. Nomura T., Watanabe M. and West T. S., *Anal. Chim. Acta.* 1985, **175**, 107.

66. Muramatsu H., Tamiya E. and Karube I., *Anal. Chim. Acta.* 1988, **60**, 2142.

67. Cady W. G., *Piezoelectricity* 1964, Publ. Dover, New York.

68. Bottom V. E., *Introduction to Quartz Crystal Unit Design*, 1982, Publ. Van Nostrand Reinhold, New York.

69. Yang M., *Ph. D. Thesis*, 1993, University of Toronto.

70. Muramatsu H., Dicks J. M., Tamiya E. and Karube I., *Anal. Chem.* 1987, **59**, 2760.

71. Lasky S. J., Buttry D. A., *ACS Symp. Ser.* 1989, **403**, 237.

72. Mc Culloch W. S. and Pitts W., *Bull. Math. Biophys.* 1943, **5**, 115.

73. Rosenblatt F., Psychological Review 1958, **65**, 386.

74. Minsky M. and Papert S., *Perceptrons*, MIT Press, Cambridge, 1969.

75. Hopfield J. J., *Proceedings of the National Academy of Sciences*, April 1982, USA, Vol. 79, p. 2554.

76. Rumelhart D. E. and McClelland J. L., *Parallel Distributed Processing*, MIT Press, Cambridge, MA, Vol. 1, 1986.

77. *Neural Computing and Technology Handbook*, NeuralWare Inc., Pittsburgh, PA, 1993.

78. Patterson D. W., *ANN Theory and applications*, Prentice Hall, Singapore, 1996.

79. McInerny M. and Dhawan A. P., *IEEE Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithm*, 1993, Innsbruck, Springer-Verlag, Wein, p. 203.

80. Hirose Y., Koichi Y. and Shimpei H., *Neural Networks*, 1991, Vol. 4, 1, p. 61-66.

81. Fahlman S. E., *An Empirical Study of Learning Speed in Backpropagation Networks*, *Carnegie Mellon Report No. CMU-CS-88-162*, 1988.

82. Matsuoka K. and Yi J., *Proceedings of the IJCNN-91*, 1991, Singapore, Vol. 2, p. 1117.

83. Van Ooyen A. and Nienhuis B., *Neural Networks*, 1992, Vol. 5, p. 465.

84. Widrow B., Winter R. G. and Baxter R. A.. *Proceedings of the First IEEE International Conference on Neural Networks*, San Diego, 1987.

85. Bain C. D., Troughton E. B., Tao Y. T., Evall J., Whitesides G. M. and Nuzzo R. G., *J. Am. Chem. Soc.* 1989, **111**, 321.

86. Yang M., Network analysis of the TSM acoustic wave sensor in the liquid phase, University of Toronto Ph.D. Thesis, 1993.

87. Yang M. and Thompson M., *Anal. Chim. Acta.* 1993, **282**, 505

88. Nuzzo R. G. and Allara D. L., *J. Am. Chem. Soc.* 1983, **105**, 4481.

89. Chambers J. Q., *Encyclopedia of Electrochemistry of the Elements*, Bard A. J. Ed., Marcel Dekker, NY, vol. XII, pg. 329.

90. Porter M. D., Bright T. B., Allara D. L. and Chidsey C. E. D., *J Am. Chem. Soc.* 1987, **109**, 3559

91. Maissel L. I. and Glang R., Eds. *Handbook of Thin Film Technology*, McGraw-Hill, NY, 1970.

92. Andrade J. D. *Surface and Interfacial Aspects of Biomedical Polymers*, Vol. 1, Plenum Press, NY, 1985.

93. Mittler-Neher S., Spinbe J., Liley M., Nelles G., Weisser M., Back R., Wenz G. and Knoll W., *Biosen. Bioelectron.* 1995, **10**, 903.

94. Furtado, M. L. and Thompson M., *Analyst* 1998, **123**, 1937.

95. Su H. and Thompson M., *Can. J. Chem.* 1996, **74**, 344.

96. Zhong Q., Innins D., Kjoller, K., Elings V. *Surf. Sci.* 1993, **290**, L688.

97. Tamayo J. and Garcia R., *Langmuir* 1996, **12**, 4430.