Generalised Sensor Linearisation and Calibration

Amra Pašić, B.Sc. (EE)

A Thesis

presented for the degree of

Master of Electronic Engineering

Supervised by Mr. Jim Dowling

School of Electronic Engineering

Dublin City University

October 2004

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Electronic Engineering (by research) is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work

Signed Anna Pasić ID 50162411

Date October 26, 2004

Abstract

The aim of this work was to conduct a survey of current sensor measurement technologies and investigate sensor linearisation, calibration and compensation methods in order to determine the methods most suitable for generic embedded sensor implementation

The thesis contains a comprehensive survey of sensor technologies and their interfacing requirements as a prerequisite for determining modules required by the generic embedded sensor interface Different linearisation and calibration techniques are investigated and the most promising techniques, curve fitting and progressive polynomial calibration method, are then examined in greater detail and simulations performed to compare their performance The fundamental limitations and trade offs in design and implementation on the microprocessor of these methods are studied The design of the compensation module is also presented and its implementation on the microprocessor in the form of the C code is described

All methods are tested and implemented on a PIC microcontroller as a part of linearisation, calibration and compensation module of the generic embedded sensor interface

Acknowledgment

I would like to thank Jim Dowling, my supervisor, for his suggestions and support during this research

I am also thankful to Claus Agersbaek, Liam Sweeney and Ciarán Watters for their guidance and help during the research They shared their knowledge with me and gave me a friendly encouragement

Thanks also to other PEI postgrads and exchange students I met during my research, Markus Florian, Eoin Kennedy, Francis Leonard, Anthony Murphy and Ralf Schlosser, for their help and friendship

Finally, I wish to thank the following Dalen (for being there for me all the time and being the best friend), Javi and Tarek (they know why), and Gabi

Amra Pašić

Contents

Chapter 1

Introduction

1.1 Background and History

Sensors produce an electrical output in response to exposure to a physical phenomenon. Processing of the output signals from different types of sensors is becoming increasingly critical as new sensor technologies are adopted and demands on signal processing, digital communications and local intelligence expand [1]. These requirements point towards the need for generic embedded sensor interfaces. The recent trend in sensor interfacing technology has been, according to Texas Instruments (TI)[2], towards the increased use of

- DSPs (Digital Signal Processors)
- FPGAs (Field-Programmable Gates)
- ASICs (Application-specific ICs)
- IGPPs (general-purpose microprocessors)

ASICs[2] can be tailored to perform specific functions extremely well, and can be made quite power efficient. However, since ASICs are not field-programmable, their functionality cannot be iteratively changed or updated while in product

development FPGAs[2] have the capability of being reconfigurable within a system, which can be a big advantage in applications that need multiple trial versions within development, offering reasonably fast time to market They also offer greater raw performance per specific operation because of the resulting dedicated logic circuit However, FPGAs are significantly more expensive and typically have much higher power dissipation than DSPs with similar functionality

In contrast to ASICs that are optimized for specific functions, general-purpose microprocessors (GPPs)[2] are best suited for performing a broad array of tasks However, for applications in which the end product must process answers in real time, or must do so while powered by consumer batteries, GPPs comparatively poor real time performance and high power consumption all but rule them out

Digital Signal Processing (DSP) provides a powerful numeric means of extracting useful information from sensor data applied to a system and it is executed by a DSP processor It is a method of processing real world signals (represented digitally) using mathematical techniques to perform transformations or extract information

It is best to optimise the use of all the resources available in any system Especially it is important to make good use of DSP and software to handle problems that are more suited to those components This refers to errors such as gain error, offset error, or non-linear characteristics of a sensor Attempts at linearisation, precise span and zero adjustments in the analogue domain are generally difficult Processor power is very inexpensive and it is often much easier to do sensor linearisation and calibration in software

Why use digital signal processing[3]?

- Flexibility - differing algorithms and coefficients may be applied at will
- Operations are noise free
- Algorithms not realizable with analogue components may be employed

- In a multi-channel system, all channel gain and phase relationships are matched

Related work in this area points to the Universal Sensor Interface Chip (USIC) which had been developed by ERA Technology Ltd. and its partners in the European Research Initiative EUREKA project JAMIE (Joint Analogue Microsystems Initiative of Europe - EU 579) in early '90s [1].

1.2 Motivation

There are innumerable sensor types which have different sensor outputs (e.g. different voltage level signals) and different needs for sensors' interfacing requirements that need to be met in order to bring the information from the real world to an observer (nowadays, more often than not a computer) via an instrumentation network (e.g. CAN bus).

Figure 1.1: Conceptual diagram.

This raises a question: How to integrate these different sensors' interfacing requirements(Figure 1.1)? The key element is the development of a single, configurable device to interface to the sensor (which converts energy from one form to another), and condition the data. Furthermore, this device must give an output in a format suitable for transmission by analogue, digital or wireless means (instrumentation network).

Looking for an answer to the question above, Enterprise Ireland (EI) funded a research project titled "Generic embedded sensor development" of which the worked described in this thesis formed a part The project was carried out by three PEI Technologies (PEI) centres the University of Limerick Analogue Centre (ULA), the University of Limerick Thermal Centre (ULT) and the Dublin City University (DCU) centre

As a part of this project, a survey of present sensor measurement technologies was undertaken and delivered in the form of a report [4] The aim of the survey was to categorise and characterise current sensor measurement techniques and interfacing requirements with a view to developing generic electronic solutions for

1 Interfacing between sensors and a local processing core,

2 Processing the data locally,

3 Interfacing between the local processing core and the next stages of the data acquisition system

The intention was for such solutions to be novel, industrially relevant and patentable

Sensor interfacing requirements are as plentiful as sensor types and sensor applications Some sensors are linear, others are non-linear, some are immune to temperature and humidity, others require compensation, some produce large voltage outputs, others require large amplification of current or voltage Some applications require analogue outputs, others serial digital links, others wireless links

Sensors produce an electrical output in response to exposure to a physical phenomenon The relationship between the sensor output (e g output voltage) and sensor input (e g gas concentration, temperature or pressure) is generally non-linear and can vary from sensor to sensor The output of the sensor is usually processed to provide a voltage that is directly proportional to the sensor input This processing is called linearisation and calibration In calibration, the sensor

is exposed to a few known inputs (e g three known gas concentrations spread over the range of a gas sensor) and the output of the sensor is noted for each input setting The performance of the sensor at these calibration points enables a linearising/calibrating function to be derived so that the output of the sensor for succeeding readings can be operated on by the linearising function to produce a voltage that is directly proportional to the sensor input (e g gas concentration)

The aims of the sub-project (part of the "Generic embedded sensor development" project) described in this thesis were to

- Determine interfacing requirements with a view to developing generic electronic interfacing solutions,
- Find the most suitable method for function approximation of calibration data,
- Model and simulate different linearisation methods, examining them in detail to find the optimal solution,
- Investigate compensation techniques,
- Validate the models against the experimental data (including iterative attempts of the model improvement)

The investigation of interfacing requirements with a view to developing generic electronic interfacing solutions for configuration, calibration, linearisation, compensation, signal conditioning and interfacing, calls for an interface that would be suitable for the vast variety of sensors on one end and different types of instrumentation networks on the other (see Figure 1 1) That interface could be called a generic embedded sensor interface (GESI)

Designing the interface between a sensor and the instrumentation network is a challenge A simplified representation of GESI is given in Figure 1 2 and consists of preamplification, digital processing (e g linearisation, calibration, compensation, filtering), interface to an instrumentation network and a user interface A user interface is the mechanism by which a user interacts with a system However

in this thesis, the user interface refers primarily to the programming behind that mechanism and it was NOT the part of this work.

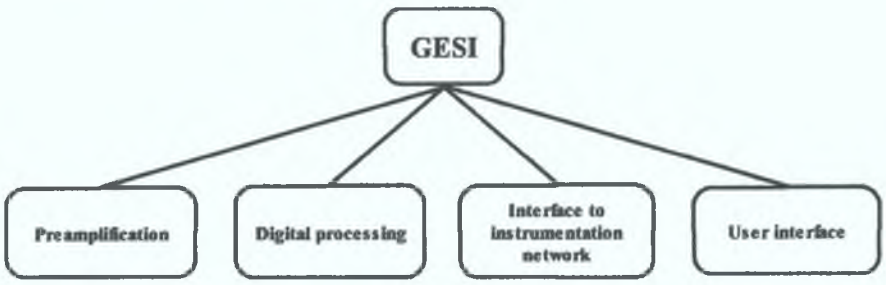

Figure 1.2: Generic embedded sensor interface.

The user interface is an important part of the generic embedded solution. It should be designed so that a user with a general technical background can configure the sensor system without programming.

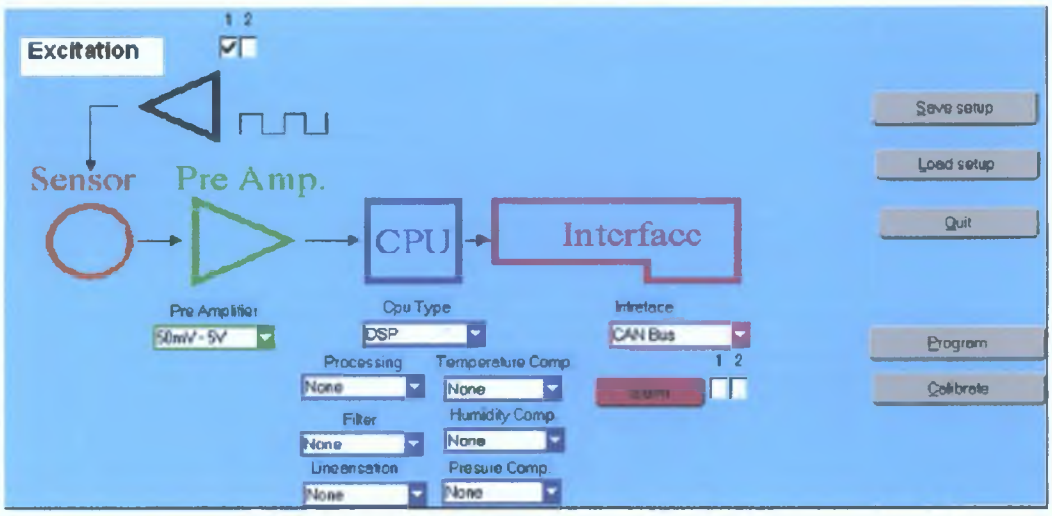

Figure 1.3: Example of GUI for generic embedded sensor interface application.

The user would select the required sensor excitation, level of sensor output preamplification, temperature/humidity compensation, signal linearisation and processing and output interface. The user interface should also facilitate calibration of the system. It is anticipated that the user interface would be implemented in the form of a Graphic User Interface (GUI) on a computer and would consist of drop-down windows (see Figure 1.3), one for each selectable parameter. The user would select the desired options from the drop-down windows and, when all

selections have been completed, the system would be configured The configured options would be effected by prewritten C code

1.3 Contributions

The main contributions of this thesis toward the aims and objectives declared above

- Sensor interfacing requirements with a view to developing generic embedded solutions were determined
- The algorithms that provide good function approximation for use in the linearisation process, based on their function behaviour investigation and estimated errors, were found to be Gauss elimination and progressive polynomial calibration method (PPC)
- There is no evidence that more calibration points in either the Lagrange interpolation or progressive polynomial calibration techniques yield better effectiveness
- The development of a compensation model for temperature and humidity

1.4 Outline of the Thesis

This thesis is divided into eight chapters as follows

Chapter 1 gives an introduction to the concept of the generic embedded sensor interface and provides a brief background and history

Chapter 2 presents a short summary of the survey of different sensor types with the emphasis on gas sensors, environmental sensors and temperature and humidity sensors It also gives a brief descriptions of communication interfaces

Their interface requirements such as sampling, antialiasing, amplification, analogue to digital conversion (ADC), calibration requirements, are investigated in Chapter 3 and the model (based on these investigations) of the generic embedded interface is proposed

Based on the aforementioned research, various linearisation and calibration techniques are examined in Chapter 4 The most promising techniques for generic sensor applications are then further investigated and simulated to compare their performances and choose the most suitable ones for generic embedded sensor interface The simulation results are shown and discussions are given in Chapter 5 including the limitations of the proposed techniques

In Chapter 6 compensation of the sensor output for temperature and humidity is explored, compensation techniques are reviewed and the design of the compensation module is offered

Tests on the prototype of the calibration and compensation module were performed and the results are presented and commented upon in Chapter 7 A description of the experimental procedure and comments on problems encountered are given The test rig setup and its associated C code description for the chosen platform (the Microchip PIC16F877 microcontroller) are included

Conclusions and recommendations for future work are given in Chapter 8

Chapter 2

Sensor Characterisation Summary and Communication Interfaces

A comprehensive survey of sensor technologies with an emphasis on potential industrial applications was carried out in order to determine the interface requirements [4] The part of the survey dealing with different sensor types is included in Appendix A and a brief summary is included in this chapter, the second part dealing with communication interfaces is also discussed here

2.1 Sensor Characterisation Summary

Sensors are examined to establish their method of operation, the excitation required, their output signal level, the analogue pre-conditioning required, their sensitivity to temperature and humidity and the signal processing required (references [5] to [6] in Appendix A) Temperature and humidity sensors, although very common and well understood, were also examined since they are often required to provide compensation for other sensors

Gas, temperature, humidity and water quality sensors together with surface acoustic waves (SAW) sensors were targeted by this survey The reason for this is the growing demand for environmental and water quality sensors and also usage of these sensors in medicine The intention was to identify challenging characteris-

tics such as linearity, compensation, calibration and input-output characteristics

There are innumerable sensor types A limited but representative number of sensor types was reviewed [4] Priority was given in the survey to gas sensors, environmental sensors, medical sensors and surface acoustic wave (SAW) sensors since there is a growing demand for each of these sensor types

Gas sensors, for example, are important for detecting, at source, the release of excessive concentrations of gases that are harmful to health or gases that contribute to global warming, ozone depletion or acid rain They are typically used in automobiles, refrigerant plant, air-conditioning plant, industrial processes and enclosed car parks

Amongst environmental sensors, ones that detect concentrations of air pollutant and water pollutant gases were examined They have to detect much lower concentrations than sensors at the source of release They are typically used in city streets, in offices, in the home and in rivers near sewage treatment plants Also some other water quality sensors have been reviewed

Medical sensors are used, for example, to monitor blood gases and pH, respiration rate and blood pressure, at rest and when exercising

Surface acoustic wave sensors are beginning to become popular in sensing a variety of environmental parameters such as humidity and pressure

The applications that were of most interest and held the greatest potential were as follows

a Semiconductor gas sensor with linearisation, temperature compensation and humidity compensation

Semiconductor gas sensors have the advantages of being cheap, robust, long life (i e do not require frequent replacement like electrochemical sensors) They have the disadvantages of being non-linear, sensitive to temperature and humidity, difficult to calibrate and responsive to several gases Their low cost and long life would mean that they would be used much more

widely if their sensitivity to temperature and humidity could be overcome or easily dealt with

b Infrared gas sensor

Infrared sensors are very selective in gas detection and very accurate but are very expensive and their outputs require significant processing If their cost could be reduced they would be used more widely

c Water oxygen content sensor using plastic film flourescence

These sensors are new and could be very cost effective They require excitation and significant signal processing

d Portable medical sensors with wireless communications

These sensors are new and would replace existing fragile wiring to patients with a very low power RF link They would be small and powered by battery Bluetooth links may not be suitable because of power and battery drain

e Surface acoustic wave sensor

Surface acoustic wave devices, in hermetically sealed packaging, have been used for many years as frequency reference or filter components in RF communication applications When the hermetic seals are removed, the devices can be configured as sensors of various environmental parameters and if absorbant surface layers are deposited they can be configured as gas sensors

The application deemed to have the greatest potential as demonstrators of the generic embedded sensor system was a semiconductor gas sensor with linearisation, compensation for temperature and humidity, analogue current loop link, and digital serial link and as such it was used for validation of different models (discussed later in this thesis) against the experimental data

2.2 Communications Interfaces

Communication methods between sensor modules and associated instrumentation and communication systems are explored Analogue methods (e g 0-10V voltage, 4-20mA current loop), digital links (e g RS232, RS485), Fieldbuses (e g Can, Lonworks), powerline communications, wireless links (e g RF and IR) and fibre optic links are some of the methods discussed Nowadays, many sensors and actuators have added intelligence which increases the use of digital communications handled by microprocessors

2.2 1 Analogue

Typical analogue output signals are those of 0-10V, 0-5V and 4-20mA

The most commonly used analog signal is 4-20mA because of the good noise immunity Current loops are virtually immune to induced electrical noise while it is difficult to shield voltage signals against noise Also 4-20mA can be transmitted over long distances without any signal degradation Voltage signals suffer losses due to the resistance of the wire Current loops are not affected as long as the total load and wire impedance does not exceed the maximum load rating of the transmitter There are different types of transmitters, 2 wire loop powered, 3 wire and 4 wire transmitters [7]

Two wire loop powered transmitter (Figure 2 1) does not require a local external or dedicated power supply The power to operate a transmitter is taken from the current loop Since there is always at least 4mA in the loop it can be used to power high impedance loads Loop powered devices have the advantage that no external power supply is required if the device consumes less than 4mA (it may be powered by the current loop)

Three wire transmitters (Figure 2 2) require an external supply, usually 24V DC Two connections are for the power supply and the third is the 4-20mA output which is referenced to the negative of the power supply Transmitter common is

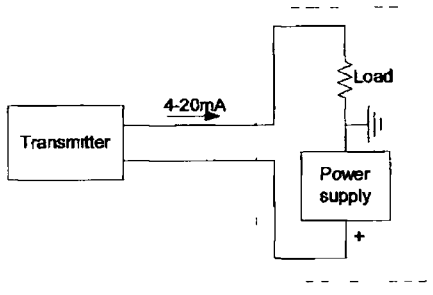

Figure 2 1 2 wire transmitter

connected to the receiver common

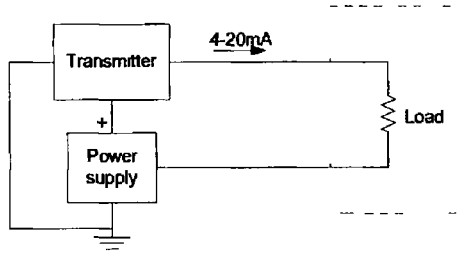

Figure 2 2 3 wire transmitter

Four wire transmitters (Figure 2 3) also require an external power supply, usually 24V DC or 120V AC The reason for the 4 wire is that some types of instruments require more power than the small amount that can be supplied by the 4-20mA signal

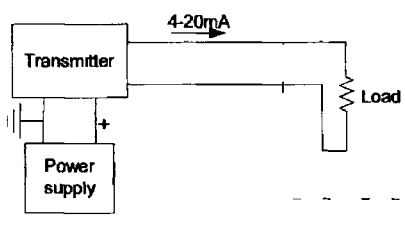

Figure 2 3 4 wire transmitter

Transmitters can be also isolated to prevent ground loops, example of an isolated 2 wire transmitter is shown in Figure 2 4

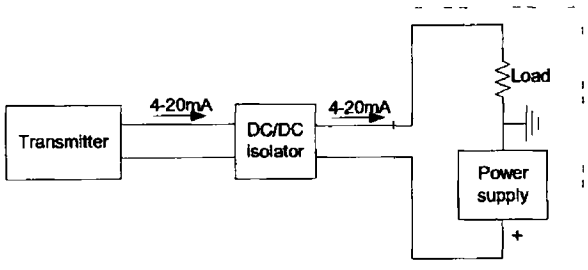

Figure 2 4 An isolated 2 wire transmitter

2 2.2 RS232, RS422 and RS485

Electronic data communications between elements will generally fall into two broad categories single-ended and differential RS232 is single-ended, independent channels are established for two-way (full-duplex) communications (Figure 2 5) The RS232 signals are represented by voltage levels with respect to a system common, there is a wire for each signal, together with the ground signal This interface is useful for point-to-point communication at slow speeds The output signal level usually swings between +12V and -12V The "dead area" between +3V and -3V is designed to "absorb" line noise In the various RS232-like definitions this dead area may vary The RS232 signal on a single line is impossible to screen effectively for noise By screening the entire cable the influence of outside noise can be reduced, but internally generated noise remains a problem As the baud rate (the number of signal level changes per second) and line length increase, the effect of capacitance between the different lines introduces serious crosstalk (this is especially true for synchronous data - because of the clock lines) until a point is reached where the data itself is unreadable Signal crosstalk can be reduced by using low capacitance cable Also, as it is the higher frequencies that are the problem, control of slew rate in the signal (i e , rounding of the sharp edges of square wave signals) also decreases the crosstalk The maximum distance will depend on the speed and noise level around the cable run [8, 9]

When communicating at high data rates, or over long distances in real world environments, single-ended methods (RS232) are often inadequate Differential

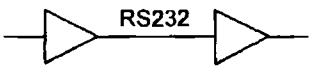

Figure 2 5 RS232 interface circuit

data transmission (balanced differential signals) offers superior performance in most applications Each signal uses one twistedpair line Both, RS422 and RS485, use the same principle, differential data transmission RS422 was designed for greater distances and higher baud rates than RS232 and is intended for point-to-point communications, like RS232 RS422 uses two separate twistedpair wires, data can be transferred in both directions simultaneously RS485 is used for multipoint communications where more devices may be connected to a single signal cable (Figure 2 6) Since RS485 uses lower impedance drivers and receivers than RS422 it can support more nodes per line RS422 can interface with up to 10 receivers for a single transmitter The limiting parameter is the receiver input impedance $R_i = 4k\Omega$ [10] 10 receivers and a termination resistor $R_t = 100\Omega$ give the maximum transmitter load For RS485 the standard (ISO 8482) specifies up to 32 drivers and 32 receivers on a single (2-wire) bus The input impedance of RS485 circuits is $R_i = 12k\Omega$ [10]

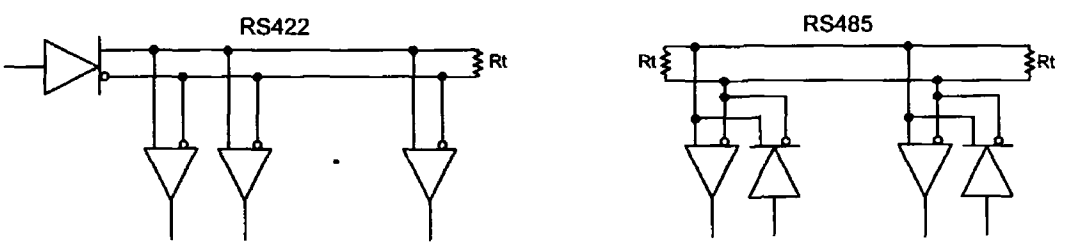

Figure 2 6 Comparison of RS422 and RS485 RS422 has one transmitter and can have up to 10 receivers RS485 has up to 32 transmitters/receivers

Since the signal is transferred via a twisted pair of wires, the voltage difference (between the wires) of common-mode interference is almost zero Differential

signals can help nullify the effects of ground shifts and induced noise signals that can appear as common mode voltages on a network as the differential inputs ignore different earth potentials of the transmitter and the receiver [11]

2.2.3 Fieldbus

Fieldbuses are all-digital, bidirectional field networks (defined by ISA standard SP50 02) intended to replace the 4-20mA standard The ability of sensors and other field devices to communicate digitally between themselves offers tremendous benefits such as reduced wiring costs, configuration time and control room space, remote diagnostics Which bus will be used depends on the details of the application The most common buses are Profibus, Lonworks, Fondation Fieldbus, CAN bus and Industrial Ethernet [12]

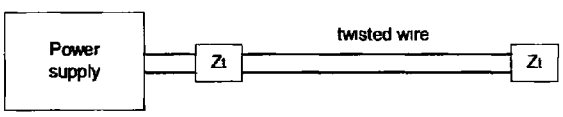

Figure 2 7 A Fieldbus

Fieldbus uses twisted pair wires A twisted pair is used, rather than a pair of parallel wires, to reduce external noise from getting onto the wires A shield over the twisted pairs reduces the noise further Fieldbus cable is similar to the type used for existing 4-20mA device wiring and, in most cases, existing cable can be used for Fieldbus When a signal travels on a cable and encounters a discontinuity, such as a wire short or open, it produces a reflection which travels in the opposite direction and is a form of noise that distorts the signal A terminator (T) is used to prevent a reflection at the ends of a Fieldbus cable (see Figure 2 7) Since the Fieldbus cable also carries power, a simple resistor cannot be used because it would use up the power intended for the devices so it requires a DC blocking capacitor $Z_t = R_t + \frac{1}{j\omega C_t}$ [13] Two terminators are needed at each network segment An ordinary power supply cannot be used because it would absorb

signals due to its low internal impedance For this reason, an ordinary power supply has to be conditioned to separate the conventional power supply from the Fieldbus wiring and this is done by putting a choke (large inductor) between the power supply and the Fieldbus wiring [13]

The length of a Fieldbus and the number of devices on a network are limited by power distribution, attenuation and signal distortion The number of devices on a Fieldbus is limited depending on the voltage of the power supply, the resistance of the cable and the amount of current drawn by each device The length of a Fieldbus network is limited by attenuation Standard Fieldbus cable has an attenuation of 3 dB/km at 39 kHz or about 70% of the original signal after 1 km If a shorter cable is used, the attenuation is less [13]

2.2.4 Powerline communications

Powerline communication (PLC) is considered in some circles to be "wireless" because of the lack of "additional" wiring (typically, the term wireless implies that information is being carried from one place to another by electromagnetic waves travelling through the air) [14] The term powerline communications refers to a form of wireless communications that employ existing AC power lines to transfer information which would normally require additional hard wire installation

Transmitting and receiving data via AC power lines is cost-effective at speeds up to about 100 kb/s (kilobits per second) [15] Powerline communication is especially attractive where walls absorb or reflect wireless systems' RF signals

Data transmission via AC power lines is very attractive because the power lines are almost everywhere in the industrialized world Unfortunately, power lines are unfriendly to data All manner of loads from switching power supplies to lamp dimmers and universal motors inject significant noise Characteristic impedance (Z_0) and attenuation as a function of frequency can change dramatically from moment to moment as various loads connect to and disconnect from the line [15] Also, the transmitted signal may reach the destination through different paths

and thus at different times causing an echo Another multipath effect occurs when some signals reach the destination as late as the subsequently transmitted signals causing interference [16]

Traditional powerline communication systems use carrier frequencies below the medium-wave (MW) broadcast band (that is, below 500 kHz) The research has shown that somewhere above the MW band, an area of the spectrum is relatively free of powerline noise, the lower frequency signals carry power, while the higher frequency signals can transmit data This frequency band must be low enough to keep the power line from exhibiting antenna behaviour [15] This radiation causes disturbances in existing communication systems A solution to this problem is to reduce the radiation, i e reduce the signal power On the other hand, reducing the signal power makes the transmission more sensitive to disturbances

With the advancements in digital communications and the advent of new algorithms for signal processing, some of the above mentioned problems have been overcome In principle, it is possible to transmit binary data at frequencies much above the 50-60 Hz used for electricity transport With frequencies of the magnitude of 150 kHz, it is possible to achieve data rates of several kb/s In the receiving unit, the low frequency power signals can be extracted out of the common signal with the help of a low-pass filter Such rates are of no practical use for modern data processing and communication systems To achieve higher transmission rates (of the order of 1 Mbps or more), it is necessary to use frequencies of the order of 10-20 MHz or even more [16]

The parameters of interest of the power lines as a data communication channel are signal attenuation, noise characteristics (including background noise, impulse noise, harmonic noise of the 50 Hz power line frequency, narrowband noise) and the phase shift caused by the power line to the transmitted signal phase The latter would be important, for example, in investigating the possibility of using M-ary phase shift keying (MPSK) and M-ary differential phase shift keying (MDPSK) for PLC

2.2.5 Wireless

Wireless communications has been widely used where it is impossible or uneconomical to run network wiring These problems can be overcome with RF technology, infrared, bluetooth, SAW and PLC (Powerline communication), where SAW and PLC are described earlier An important property of signals at RF, and particularly at higher microwave frequencies, is their great capacity to carry information RF transmitters/receivers do not require a line-of-sight connection the way IR transmitters/receivers do [17]

Two principal types of infrared exist direct and diffuse [18] Direct point-to-point IR requires pointing one device at another Diffuse IR does not require a direct line of sight from one device to another (it relies on reflections from the ceiling or other reflectors), but it will not pass through opaque/solid obstacles Nowadays the best choice of wavelength band for the majority of wireless infrared applications is the band 780-950 nm because of two reasons First, in this band low-cost Light-Emitting-Diodes (LED's) and Laser-Diodes (LD's) can implement infrared transmitters Second, cheap low-capacitance silicon photodiodes operate at their maximum response in this wavelength band However, the radiation in this band can be harmful to the human eyes which leads to the use of shorter wavelength bands [19]

A great advantage of IR over RF, for in-building communications, is the low risk of health hazards (possible danger of cancer) for individuals working or living in a building who would otherwise be constantly bathed in RF radiation With IR there is no indication of such a risk

Bluetooth is a recent wireless technology currently employed in telecommunications, however one can apply the same principles to the data links between sensors and electronics without thinking of the wires and cabling between the sensors and the electronics It gives freedom to place the sensors at the right position and not where construction allows Bluetooth is a universal radio interface in the 2 45 GHz frequency band (Industrial, Scientific, Medical - ISM band) that

enables portable electronic devices to connect and communicate wirelessly via short-range, ad hoc networks Each unit can simultaneously communicate with up to seven other units It has a raw data rate of up to 1Mb/s [20]

A tiny Bluetooth microchip, incorporating a radio transceiver, is built into digital devices [21] The normal output power is very low, only 1 mW, which gives a working range of about 10m [22] It facilitates fast and secure transmission of both speech and data, even when devices are not within line of sight [21] It's predicted that by 2005, Bluetooth will be a built-in feature in more than 670 million products on the market, from communications devices to headsets and portable PCs to desktop computers and notebooks [23]

2 2.6 Fiber optics

Fiber optic transmission systems - a fiber optic transmitter and receiver, connected by fiber optic cable - offer a wide range of benefits not offered by traditional copper wire or coaxial cable [24]

1 The ability to carry much more information and deliver it with greater fidelity than either copper wire or coaxial cable

2 Fiber optic cable can support much higher data rates, and at greater distances, than coaxial cable, making it ideal for transmission of serial digital data

3 The fiber is totally immune to virtually all kinds of interference, including lightning, and will not conduct electricity It can therefore come in direct contact with high voltage electrical equipment and power lines It will also not create ground loops of any kind

4 Transmission of signals over fiber optic cable also significantly reduces the incidence of RF interference caused by the use of copper cabling

5 As the basic fiber is made of glass, it will not corrode and is unaffected by most chemicals It can be buried directly in most kinds of soil or exposed to

most corrosive atmospheres in chemical plants without significant concern

6 Since the only carrier in the fiber is light, there is no possibility of a spark from a broken fiber

Even in the most explosive of atmospheres, there is no fire hazard, and no danger of electrical shock to personnel repairing broken fibers

A fiber optic transmission system is more expensive than copper transmission system It can range in price from 60 to several thousand euros

2.3 Summary

It was decided to review a limited but representative number of sensor types in the survey (the major part of that work is in Appendix A) Precedence was given to gas sensors, environmental sensors and surface acoustic wave sensors since there is growing demand for all of these sensor types Gas sensors are important for detecting, at source, the release of excessive concentrations of gases that are harmful to health or gases that contribute to global warming, ozone depletion or acid rain They would typically be used in automobiles, refrigeration plants, industrial processes and tunnels Environmental sensors detect concentrations of air pollutant and water pollutant gases They would have to detect much lower concentrations than sensors at the source of the release They would typically be used in city streets, in offices, in the home and in rivers near sewage treatment plants Medical sensors are used to monitor blood gases and pH, respiration rate and blood pressure when exercising Surface acoustic wave sensors are beginning to become popular in sensing a variety of environmental parameters such as humidity and pressure

In addition to the sensors survey, most common sensor communications interfaces are considered and briefly described

Chapter 3

Sensor Interface Requirements

This chapter discusses interfacing requirements between the sensor and the instrumentation network with a view to developing generic electronic interfacing solutions A simplified block diagram of these requirements is given in Figure 3 1

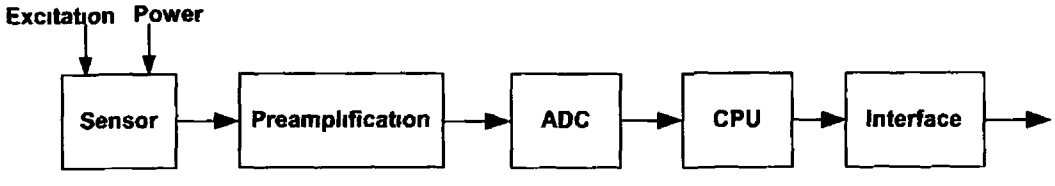

Figure 3 1 A simplified block diagram of the sensor interface

Some sensors such as gas semiconductor sensors, need excitation or heating The sensor output is often a very small analogue signal and issues such as amplification and analogue to digital conversion (ADC) are discussed in this chapter Calibration requirements, compensation and linearisation of sensor modules are also things to consider The linearisation methods will be discussed later on in more detail in the thesis

Also, general system requirements are examined Topics discussed include power supply and signal isolation, battery and intrinsically safe operations

3.1 Sensor Interfacing Requirements

There are many different sensor types The previous chapter and Appendix A deal with various types of gas sensors, temperature sensors, humidity sensors, different types of water quality sensors and sensors based on Surface Acoustic Waves (SAW), which interfacing requirements are used as a starting point for the generic embedded sensor interface requirements

Some sensors need excitation such as a type of SAW sensor or heating like semiconductor gas sensors Most sensors need power Each sensor can have its own power converter or there can be a central converter with DC power bussed to the individual sensor

Often the raw signals must be amplified, smoothed and sometimes translated Typically then one or more operational amplifiers may be used before the ADC, each of these will contribute some amount of analogue error to the digitized data, as will the ADC itself Even small errors will use up several bits of resolution 10 bits of ADC conversion resolution, over a 5V range, makes each bit worth 5mV This implies the need for high accuracy systems, those with a 10 or more bit ADC resolution

The output of a sensor, a very small voltage or current, can be amplified with low offset voltage, low bias current amplifiers The ideal voltage and current operational amplifier (op amp) inputs do not draw any currents ($Ip = 0$ and $I_n = 0$),the voltage difference between the inputs is zero ($v_p = v_n$), infinite open loop gain $A \rightarrow \infty$ and infinite bandwidth (Figure 3 2)

Non-ideal behaviour of operational amplifiers is characterised by input offset voltage, input bias currents, thermal drift, finite common-mode rejection ratio (CMRR), finite power-supply rejection ration (PSRR) [25]

The amplified analogue signal can be transmitted in analogue form over wires to the other parts of the instrumentation network or more frequently analogue inputs are converted to digital signals by an ADC Analogue to digital conversion is limited to a small range of voltage, so that internal or external input circuitry

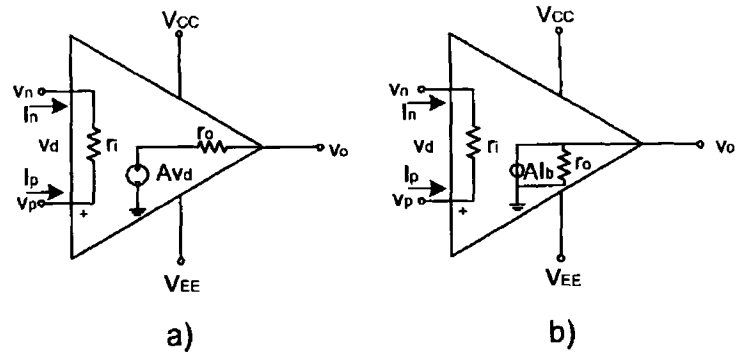

Figure 3 2 a) Voltage amplifier, b) Current amplifier

must change the character of non-compatible signal types to a voltage range within the limits of the ADC, i e if the maximum operating voltage of ADC is 5V, and the signal that has to be converted with the ADC has a 10V range, then the 10V signal can be attenuated to a 5V range with a voltage divider for instance Conversion time (time needed to produce a digital value corresponding to the analogue value) of ADCs usually depends on the resolution (which is directly related to the accuracy) and the clock frequency and it has to be taken into account when choosing the ADC

The ADC must have a steady signal in order to accurately perform a conversion, thus it is often preceded by a sample and hold (S/H) circuit (Figure 3 3)

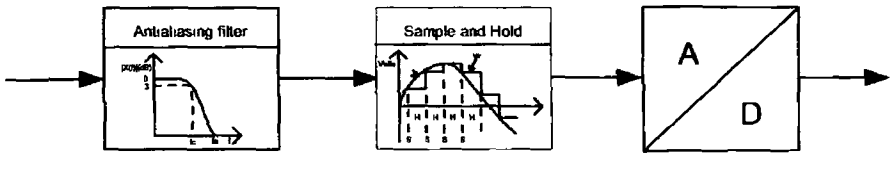

Figure 3 3 Data acquisition block diagram for ADC

The ADC may also need an antialiasing filter on its input (see Figure 3 3) The output from the ADC is highly dependent on the ADC sampling frequency and the maximum frequency of the analogue input signal [26] A low-pass or band-pass antialiasing filter placed immediately before the ADC band-limits the analogue input This ensures that the original input signal can be reconstructed exactly from

the ADC's output samples To achieve this, the minimum sampling frequency should be equal to the Nyquist frequency f_N which is twice the bandwidth of the analogue signal $2f$ Undesirable signals above $f/2$ can create spectrum overlap and add distortion to the desired baseband signal Oversampling frequencies (sampling rate greater than the Nyquist rate) can be also employed, drastically reducing the steepness of the antialiasing filter rolloff However the oversampling technique requires fast ADCs [26]

Many sensors have outputs that are nonlinear and that are sensitive to temperature and humidity In these cases linearisation, temperature compensation and humidity compensation are required Linearisation, temperature and humidity compensation can be performed by a microprocessor But even in a digital signal processing environment an analogue calibration of the sensors and compensation for temperature effects or nonlinearities often make sense before the signal is applied to ADC Compensation of offsets results in maximum signal to noise ratio The offset compensated signal allows the use of the maximum gain of an amplifier or the full input range of an ADC thus minimizing the influence of noise and distortions

3.2 Calibration Issues

Many sensors are used in automated systems in industrial plants They need to be calibrated when first installed and they need recalibration as they drift with age and process conditions (such as temperature and pressure) After some change is made to a new instrument or when old sensors are being replaced with new sensors, a recalibration is needed in order to compare the sensor signal with previous measurements of the reference data base

In practice, the sensor calibration is performed by putting the sensor in a controlled environment [27] The environment is controlled in such a way that the physical reference signals (reference values for the measurand) can be generated in accordance with official standards The corresponding sensor output signals

are then recorded The sensor outputs are then processed in order to retrieve a signal which relates to the original physical signal in accordance with the desired transfer characteristic Sometimes, more (reference) sensors are needed to monitor more ambient parameters (e g temperature and humidity) The measurements of the sensor output signal at several temperatures can be used to determine the error with respect to the desired sensor transfer characteristic and therefore allow for compensation to be effected One of the weakest points of any system is the analogue sub-system drift, aging, and even humidity and cleanliness can dramatically affect the response of an analogue circuit Traditional analogue circuits use potentiometers to remove most errors But even employing only two potentiometers (offset and gain correction) can lead to difficulties as setting the gain potentiometer often causes readjustment of the offset one In the case of a larger number of potentiometers this becomes an iterative process and it is not suitable for generic applications since they must support different kinds of sensors with different calibration requirements [28]

A microprocessor can be used for interpreting complex calibration curves and presenting digital readouts Embedded systems can take advantage of the processor power to mathematically remove many of the error sources, including the errors of the ADC itself, because digital calibration affects the transfer function of the entire system (overall transfer function) This can increase overall system accuracy, since the processor can perform recalibration continuously In the real world potentiometers will only be adjusted infrequently, if at all They should be eliminated and the system should be self calibrating Even the simplest of analogue circuits are subject to offset and gain errors A self calibrating system needs some known (reference) value which is injected into the front-end during the calibration and an algorithm to correct the overall transfer function to meet these known values [28]

Most analogue circuits are linear If the system exploits linear sensors (or at least sensors that are mostly linear in the operating range) and linear circuits, it follows that use of a linear correction will remove essentially all of the error Even some non-linear circuits can use a linear correction if the operating range is

sufficiently restricted To solve the linear equation $y = mx + b$ for two unknowns m and b, we need at least two equations and this is easily performed on the microprocessor [28]

Unfortunately many sensors are non-linear and they need more calibrating points If a sensor is linear and not sensitive to temperature or humidity one or two point calibration is needed Two, three or four point calibration is needed if it is linear but sensitive to temperature or humidity, or non linear and not sensitive to temperature and humidity Non linear and sensitive to temperature and humidity needs even more calibrating points, because of the temperature and humidity compensation

Also a noisy system can produce bad calibrations This can be overcome by averaging the readings first, and then performing the computations

Drifty sensors will make the calibration procedure more complicated It's hard to create a standard physical parameter (like temperature or a light level) with precision This makes sensor calibration difficult, but such calibration is needed regardless of the instrument's technology

3.2.1 Calibration requirements of sensors covered by Appendix A

Since the response of gas sensors may not be stable in time they require a frequent calibration Gas-sensor properties that may be affected by time-dependent drift include the offset, the signal to noise ratio and the sensitivity

Relative humidity (RH) sensors are unreliable and they must be calibrated on installation and about every six months thereafter Calibration involves checking the accuracy at least two RH values

Most temperature sensors need calibration Fiber optic and thermocouples temperature sensors do not need recalibration

The calibration interval needed for pH/ORP (where pH measures the concentration of hydrogen ions in the water and ORP stands for oxidation reduction potential - described in Appendix A) is highly application-dependent The more uniform process conditions are (for instance, temperature, pressure, composition or freedom from coating), the more stable a pH sensor will be and the longer the calibration interval may be Most pH installations calibrate between once a week and once a month but shorter or longer intervals may be appropriate based on experience

Conductivity and dissolved oxygen sensors also need calibration and the calibration interval is application-dependent

Every sensor does have its unique calibration requirements This can be realised in an inexpensive way (less time consuming) using automated calibration procedures The cost can be further reduced by optimising the number of the calibration measurements for the particular sensor

Calibration of instruments and measuring devices is of paramount importance in order to obtain accurate, precise data from which decisions can be made These decisions will most likely have profound effects on the quality and quantity of products produced, the validity and significance of research and development, or the justification of a health alert or drug recall [29]

3.3 System Requirements

3 3 1 Isolation of power supply and signal

Isolated supplies preserve device accuracy by preventing ground loops, and they easily provide positive regulated voltages from a negative power bus without compromising the benefits of that bus [30]

Signal isolation enables digital or analogue signals to transmit without galvanic connections between the transmitting and receiving side of the isolator This in

turn allows ground or reference levels on the transmitting and the receiving side to differ by up to thousands of volts, and prevents circulating currents between differing ground potentials that can contaminate signals The noise on the ground signal may interfere

In some applications, a galvanic connection between reference levels may provide a current path that would present a safety risk to an operator or medical patient [31] Analogue signal isolation is not just for preventing injury and damage from high voltages but can also reduce the noise attributable to common-mode voltage [32]

Some analogue isolators use optoisolation techniques Despite their popularity with digital signals, whose precise waveshapes are usually unimportant, optoisolators play only a minor role in the analogue signal isolation market Most analogue isolation circuits use transformers Transformers inherently provide ohmic isolation Unfortunately, most analogue isolation applications involve sending DC across the isolation barrier, and transformers work only for AC Therefore high frequency (HF) modulation techniques have to be applied [32]

The isolated transmission of digital signals has been widely used in computer automatic monitoring It can eliminate the mutual interference and achieve more efficient and reliable data transmission There are mainly two kinds of isolated transmission techniques One is pulse transformer isolation, the other is optoelectronic isolated transfer [33] The advantage of pulse transformer isolated transmission is that the transmitting distance is long, up to 100m, the transmitting rate is fast (from 500Kbps to 5Mbps) and the transmission has the function of parity check Optoisolated communication is often used in multiplex system with bit serial transmission For multiplex distributed systems without a very high speed (up to 115 2Kbps), the optoelectronic isolation communication is a simple and cheap solution [33]

3 3.2 Battery operated

In some applications power supplies cannot be used, e g portable sensors, medical sensors worn during exercises, and battery power must be used To be battery operated a sensor needs low power consumption and its supply voltage should match the battery voltage

3.3.3 Intrinsically safe operation

Intrinsic safety (not considered further in the thesis) is a protection concept employed in potentially explosive atmospheres It is suitable only for low current apparatus such as instrumentation and control circuits Limitations are placed on current and voltage to ensure that there is insufficient energy in the circuit to ignite gases/vapours (intrinsic safety protection type BS5501 Part 7 EN50020), by either thermal or electrical means [34] In order for ignition to occur, a certain amount of energy is needed The minimum ignition energy is defined as the smallest possible amount of energy which is converted during the discharge of a capacitor and is just enough to ignite the most ignitable mixture

In electrical circuits the mechanism for the release of this ignition energy is one or more of the following [35]

- Open circuit or short circuit components or interconnections in a resistive circuit,
- Short circuit of components or interconnections in a capacitive circuit,
- Open circuit components or interconnections in an inductive circuit,
- Ignition by hot surfaces

In a circuit that is non-reactive (quasi non-inductive/capacative) there is no stored energy to be released in an arc The main consideration, therefore, is the amount of energy in the circuit The power provided to intrinsically safe

equipment is normally via zener diode barriers or galvanic isolators Barriers are widely used in intrinsic safety to allow connection to power supply lines Zener diodes limit the voltage to the hazardous area by shunting excess energy to an equipotential ground system They are used to limit both voltage and current (and hence power) into the hazardous area of the equipment If a circuit can be proven to have no means of dissipating stored energy, and the voltage and current to it were proven to be safe, then the circuit will be safe electrically [35]

Stored energy in capacitors that could be released as a spark is an obvious problem (as the energy by-passes the resistive safety protection of the supply) The worst case condition is if all the circuit capacitance was connected in parallel at worst case tolerance

In a practical circuit it may be found that when all components are added in their worst case combinations the circuit is found to be ignition capable A circuit can be made safe by employing various design techniques to limit the permutations of component connection [35] Inductive circuits are also calculated to give the worst scenario

As a component with a high surface temperature could cause an ignition, the maximum temperature that any component on a circuit board under fault conditions must be considered Smaller components will need to reach a higher temperature than the auto ignition temperature of a gas to cause an ignition

3.4 A Short Summary of Papers on Generic Sensor Electronics

1 *IEE Colloquium on Advances in Sensors, 1995, Applications of a Universal Sensor Interface Chip (USIC) for Intelligent Sensor Applications, Wilson P D , Spraggs R S , Hopkins S P , Lewis I , Skarda V , Goodey J*

Questionnaires from over three hundred companies were returned to ERA from European sensor users and manufactures Main findings of the survey

are the following functional requirements measurement accuracy from 8 to 20 bit is required for voltage, current, resistance, capacitance and frequency based sensor inputs, 70% of the applications require at least one compensating or reference input, strong interest was shown in signal conditioning with an average of 70% of the responding organisations having interest in these functions offset, span, linearization, auto calibration and self test, both analogue and digital outputs from the sensors are required, both voltage and current loop outputs are in stronger demand than any one digital output, size and power consumption are also important factors

To meet a good crossection of requirements, a Universal sensor interface module,proposed in this paper, should be a multichip module consisting of a set of common modules sensor adapter interface, DSP, associated memory module and RF transceiver module The sensor adapter interface would have ADC to perform analogue to digital conversion Analogue multiplexers would allow up to three input sources for each converter Chopper stabilised op-amps would have many uses including active analogue filters, current sources, integrators or signal summing Analogue outputs are generated by DAC which can be also put to other uses in cases where analogue input has a significant DC offset for example, to reduce it It could also be used in conjunction with an internal digital sine wave generator to stimulate sensors and bridge devices

Calibration, linearization and offset correction would be handled by the processor

2 *ESSCIRC'99 Proceedings of the 25th European Sol , 1999, A Fully Integrated Sensor Interface Chip, McCarteney D, et al*

The paper describes a chip that contains two 20-bit sigma-delta ADC The main ADC has a rail-to-rail input buffer and programmable gain amplifier (PGA) There is a 12-bit DAC and current sources for sensor excitation and circuitry for sensor fault detection An 8052 microcontroller core is provided for processing the ADC outputs There is 640 bytes of user flash

memory and 8k bytes of program flash Clocking is from a 32 kHz watch crystal oscillator An on-chip PLL multiplies up this frequency to provide programmable clocks for the microcontroller

Also there is a need to integrate the digital signal processing An emerging standard, IEEE 1451 2 has partioned the sensor interfacing system into two aspects the measurement aspect is dealt with by the Smart Interface Module (STIM) while the application aspects are handled by the Network Capable Application sensor (NCAP)

3 *Proceedings of the SPIE, 1999, Universal Sensor Interface Module, King D , Torres A , Wynn J*

A low power Universal Sensor Interface Module (USIM) is being developed by Raytheon-TI Systems Company for use with fields of unattended distributed ground sensors This multi-chip module consists of a sensor adapter interface, DSP and RF transceiver module The sensor adapter interface is designed for video and IR sensors and non-imaging sensors For the DSP module, a Motorola MPC555 is chosen due to its characteristics The DSP module is low power design, supports a variable clock rate such that the processing power is a function of the processing load, has the double precision floating point, good software development tools and a large memory

4 *Sensors and Actuators A (Physical), 1996, The Future of Sensor Interface Electronics, Yamasaki H*

The Future of Sensor Interface Electronics" is about future technologies on image processing The functions of interface electronics are expanding in several directions

- extension towards multidimensional sensing systems,
- extension towards multilayer signal processing systems,
- extension towards multimodal sensing systems,

- progress towards more precise sensing systems,
- progress towards more reliable sensing systems,
- progress towards more human-friendly sensing systems,
- progress towards more compact sensing systems

3.5 Summary and Conclusions

Sensor interfacing requirements are as plentiful as sensor types and sensor applications Some sensors are linear, others are non-linear, some are immune to temperature and humidity, others require compensation, some produce large voltage outputs, others require large amplification of current or voltage Some applications require analogue outputs, others serial digital links, others wireless links

The generic embedded hardware and software should facilitate the following sensor interface requirements

a Excitation of the sensor

b Amplification with precision low offset voltage amplifiers,

c Analogue to digital conversion with 10 bit minimum resolution,

d Temperature and humidity compensation,

e Response linearisation,

f Calibration,

g Self test,

h Digital signal processing,

i Analogue current loop (4-20mA) interface,

j Serial digital interface e g RS232, RS48,

k IR communication

A block diagram of a generic embedded system is shown in Figure 3 4

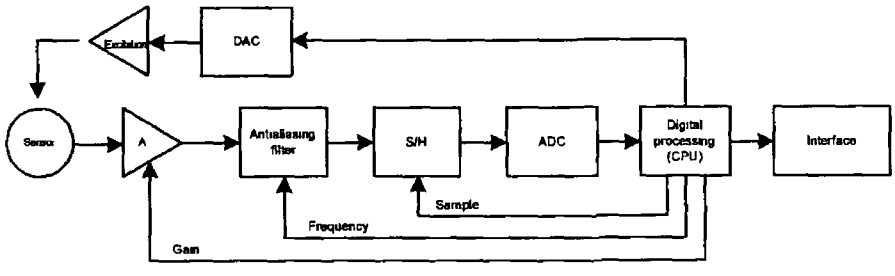

Figure 3 4 A block diagram of a generic embedded sensor interface

The application deemed to have the greatest potential as demonstrators of the generic embedded sensor system is semiconductor gas sensor with linearisation, compensation for temperature and humidity, analogue current loop link, and digital serial link The reason of choosing this sensor is its highly non-linear characteristic and need for temperature and humidity compensation

Chapter 4

Function Approximation of Calibration Data - Theoretical Background

This chapter deals with sensors that produce an electrical output in response to exposure to physical phenomenon A sensor output is (usually) intended to be uniquely related to its input In this way, from a measurement of the sensor output, the input is then "known" To insure that this is so the sensor has to be calibrated before use

Furthermore, the relationship between a sensor's output (e g output voltage) and sensor's input (e g gas concentration, temperature or pressure) is generally non-linear For many reasons (ready availability of instrumentation and use in control systems, for example) it is desirable that the signal representing the sensor input (e g gas concentration) is linearly related to that input To achieve this the actual sensor output must in general be linearised The combined calibration/linearisation process is illustrated in Figure 4 1

The output of the sensor v_{raw} (Figure 4 1) can be unknown or the manufacturer can give some specifications (data sheets) that can vary in quality In calibration, a sensor is exposed to a few known inputs x (e g three known gas concentrations spread over the range of the sensor) and the output of the sensor is noted for each input setting $v_{raw} = f_s(x)$ Processing of v_{raw} is performed in such a way as to provide a desired corrected sensor output $v_{corr} = f(x)$ which may be simply an

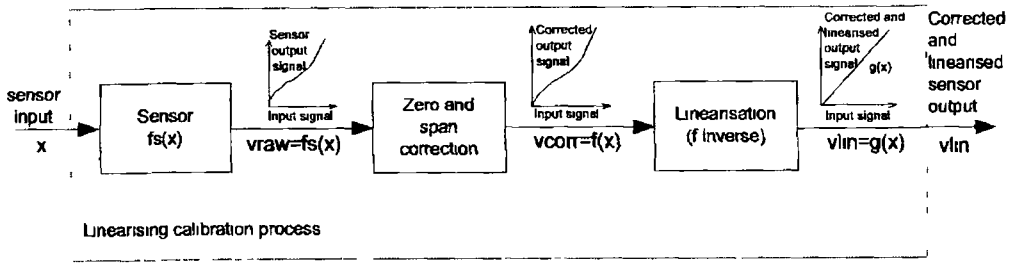

Figure 4 1 Linearising calibration process

adjustment of maximum and minimum points, normally called "zero" ("zero" is not always represented by actual zero voltage, it could be also a "live" zero) and span, and not concerned about the "shape" or "form" of $f(x)$

Then if, as is often the case, a linearised output signal is desired (i e a final desired output $g(x)$ that is linear), a process of linearisation $v_{lin}(x) = g(x) = ax+b$ with, $(v_{lin})_{max} = g(x_{max})$ and $(v_{lin})_{min} = g(x_{min})$ is required a is simply a scaling factor and b is the desired intercept, often equal to zero

Without loss of generality one can consider $a = 1$ and $b = 0$ and thus, $g(x) = x$ Then, linearisation is basically finding the "best" approximation function so $x = f^{-1}(v_{corr})$ and the overall process could then be called the linearising calibration process For example, a gas sensor desired output could be specified as 1V for the minimum gas concentration and 5V for maximum gas concentration at the sensor input with intermediate final outputs linearly related to the input The behaviour of the sensor at a specified number of points would then enable a linearising calibration function to be derived If a system can be designed to perform linearising calibration automatically then such a system can be subsequently used to correct (zero and span) and linearise the actual sensor output when the sensor is in use (e g laboratory, industrial plant or "in the field")

In this chapter, various linearising calibration techniques are initially investigated The most promising techniques for generic sensor applications (i e curve fitting and progressive polynomial calibration method as described later) are then examined in greater detail and simulations are performed to compare their performance Two of the curve fitting methods and the progressive polynomial method

are further investigated regarding their suitability for implementation on a microcontroller

4.1 Introduction

If sensor data are to be subsequently processed digitally and a sensor output is nonlinear the simplest way to perform linearisation is in the digital domain Even though, in a digital signal processing environment, an analogue calibration of the sensors and compensation for temperature effects or nonlinearities often make sense before the signal is applied to an analogue to digital converter (ADC), it is better to preform it digitally This way, besides errors in the transfer function of the sensor, errors from the other function blocks, such as the ADC or some analogue interface, are corrected automatically in the calibration process

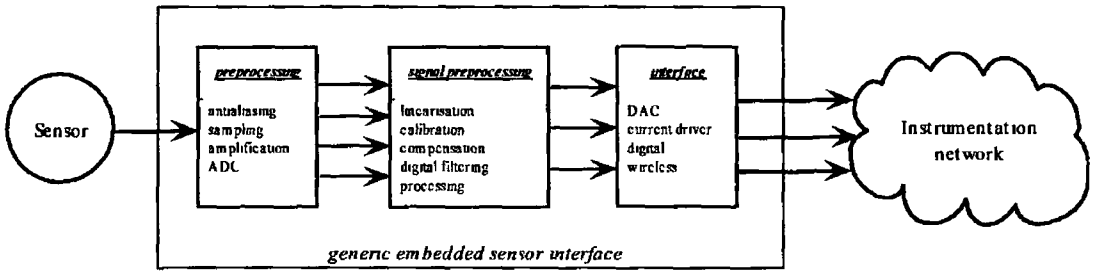

Figure 4 2 GESI structure

Since the calibration requires the computational power of the system and it is useful that the calibration information is not attached to the system's computer but the sensor itself, microcontrollers play an important role in GESI A structure of a GESI (introduced earlier in Chapter 1) is shown in Figure 4 2 Since a microcontroller affects the cost of the GESI system and performing linearisation and calibration is power consuming, it is necessary to find the trade off between linearisation and calibration techniques and processing power

4.2 Errors in the Sensor Transfer Curve

A sensor transfer function can have one or more of the following error types [27]

- Offset
- Gain, range or full-scale error
- Nonlinearity
- Cross-sensitivity
- Hysteresis
- Drift

All of them can be compensated or calibrated except drift errors which are difficult or impossible to compensate or calibrate, as this would require prediction of future errors and time measurement

4.3 Linearising Calibration Methods

A certain number of measurements need to be taken in order to determine and correct a sensor's nonlinearity The number of measurements necessary to reduce the linearity error depends on the linearising calibration method used and to reduce the costs of calibration it is important to minimize the number of measurements Costs refer to the expense of processing power and time This is an important criterion in the selection of an appropriate linearising calibration method for sensor calibration

All of the following methods [27] are based on the use of calibration measurements

- Look-up table,
- Piecewise linear interpolation,
- Piecewise polynomial or spline interpolation,
- Error minimization,

- Sensor characteristic linearisation,
- Curve fitting,
- Progressive polynomial calibration

A brief description of each of these is given below after which then follows more detailed approach to methods which were chosen for further investigation

Look-up table

In this method, the sensor output is measured for a large number of known input signals and the values, so obtained, are stored in a table and used as the ideal inverse sensor transfer function (Figure 4 3) For a particular sensor output, the corrected sensor output value can be obtained by looking it up in the table During calibration, an ADC is used to bring the sensor output signal into a digital form and this can be used as an address for the memory location holding the corrected digital value for the ideal transfer curve The advantages of this method are that calibration can be realised in a simple manner and, once the look-up table is generated, the sensor signal correction is fast and does not require any signal processing The disadvantages are that a large memory is required and a large number of calibration measurements are needed

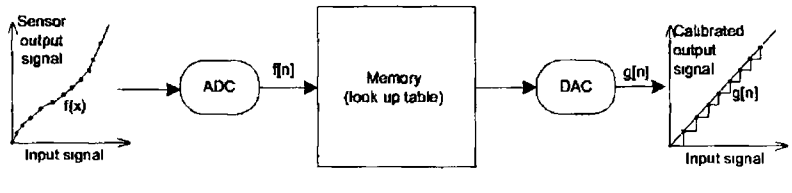

Figure 4 3 Look-up table calibration method

Piecewise linear (PWL) interpolation

In this method, the sensor output is measured for a small set of known input signals These calibration points become anchor points or "break" points on the

ideal inverse transfer curve The inverse transfer curve is completed by joining the "break" points with straight lines $y = a_n + b_n x$ These lines form a piecewise linear interpolation of the ideal inverse sensor transfer curve (Figure 4 4) The parameters that need to be stored in memory are the subrange limits and the required offset and gain corrections for each subrange The advantages are a small number of calibration points, a small memory and signal processing that is relatively simple The disadvantages are that error correction is modest, a large number of calibration points are needed for very nonlinear transfer curves and separate calculations of offset and gain for each subsection are needed

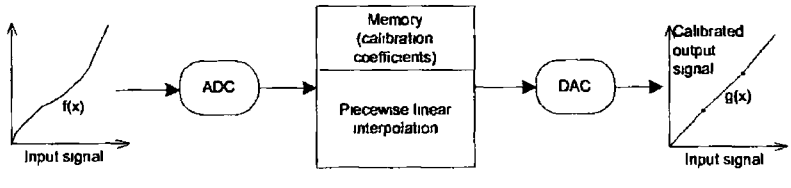

Figure 4 4 Piecewise linear calibration method

Piecewise polynomial or spline interpolation

The sensor output is measured for a small set of known input signals These calibration points become anchor points or "break" points on the ideal inverse transfer curve The transfer curve is completed by adding polynomial curves ($y = P_n(x)$) between the "break" points where these curves are derived by polynomial interpolation between three or more "break" points These interpolations can also be expressed mathematically as splines These interpolations or splines form a piecewise polynomial or spline function representing the ideal inverse transfer curve over the complete input range The advantages are fewer calibration points than PWL, good linearity that can be obtained with few calibration points and recursive equations suitable for software implementation can be used The disadvantages are that the method requires advanced computations and storing different correction functions with different coefficients for each subsection of the transfer curve

Error minimization

The error minimization is based on constructing the curve that does not exactly go through calibration measurements The coefficients of the approximation curve are adjusted in that way that the errors at the calibration points are minimized Usually the mean square error ("Least Squares") is taken as the error function that should be minimized The advantages are that good linearisation can be achieved, but this is only the case in the case of a large number of calibration measurements The storage of many calibration measurements is needed because they are used in the error function optimization In addition, there are problems of finding the optimum order for the linearisation curve (number of coefficients), determining a good error function and how to minimize it Usually the choice of a linearisation function is based on some "knowledge" about the function These are complex issues, especially for high order linearisation and *nth* order calibration (compensation) which makes this method not very suitable for realization In addition to these disadvantages this method employs complex calculations of calibration coefficients

Sensor characteristic linearisation

Sensor characteristic linearisation is based on function transformation (Figure 4 5) but it is inappropriate when it comes to generic methods as it requires *a priori* knowledge of a sensor characteristic

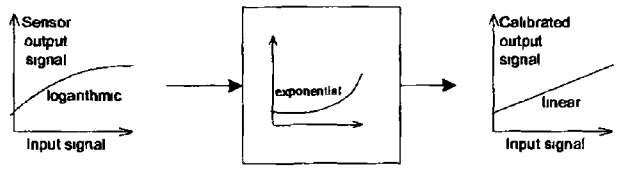

Figure 4 5 Sensor characteristic linearisation

If, for example a sensor transfer function is described by $v = ax^b$ then it is possible to determine a and b (for example by transforming the expression to $V = bX + n$

where $V = \log(v)$, $X = \log x$ and $n = \log a$ and using "Least Squares") Knowing a and b, the inverse function can be easily determined ($x = \sqrt[1/b]{\frac{v}{a}}$) in order to linearise the sensor transfer function This method is described later on in Subsection 4 4 5 The only advantage is that it is relatively easy to do it in software The main drawback of the method is that the nonlinearity has to be well known The other disadvantage is that the transformation of the nonlinear function is usually not good enough to reduce the nonlinearity

Curve fitting

When the sensor output is measured for a small set of known input signals an interpolation algorithm is then used to compute a sensor transfer function using curve fitting techniques (Figure 4 6) The inverse transfer function is then calculated so that input signal can be derived from the sensor output The other possibility is to find a curve fit function for the inverse sensor transfer function straight away The advantages of curve fitting are an inverse transfer function is obtained for the complete signal range and memory requirements are low due to a small number of the coefficients The disadvantages can be that sometimes higher order polynomials are necessary to get the desired accuracy This involves more complex calculations and may require high accuracy computations (floating point arithmetic) as well

Progressive polynomial calibration

Progressive polynomial calibration operates on the principle that each calibration measurement is used directly to calculate one calibration coefficient in the correction function This correction is then applied to the sensor output Each step is independent of the previous one The first measurement is used to correct the offset, the second corrects the gain and all the rest of the measurements are used for nonlinearity correction The advantages are that it gives good linearisation for a minimum number of points, has low memory requirements due to a small number of the coefficients and uses a repetitive algorithm (step by step calibration)

The disadvantages are that some knowledge is needed when choosing calibration measurements

4.3.1 Summary

Look-up table, error minimization and piecewise polynomial or spline linearisation need many calibration measurements and they require large memory for storage Piecewise linear interpolation may not be good enough for nonlinearity correction Curve-fitting and progressive polynomial calibration offer the possibility to linearise a sensor transfer function, using a low number of calibration measurements Hence, *curve fitting* and *progressive polynomial calibration* are chosen for further investigations due to their advantages, and discussed in the following sections

4.4 Linearisation and Calibration Based on Curve Fitting

This method consists of the calculation of the correct values for calibration coefficients and the calculation of the linearised output signal based on the actual sensor output and the calibration coefficients[27] Given that the input signal x and output signal v are both available, there are two ways of doing it

- Estimating the transfer function $v = f_{est}(x)$ and then deriving an inverse transfer function f_{est}^{-1} (notation means $(f_{est})^{-1}$),
- Composing a curve fit function for the inverse sensor transfer function right away $x = f_{est}^{-1}(v)$ (notation means $(f^{-1})_{est}$)

The latter is much easier to implement and involves matching the resulting transfer function $h(x) = Hf(x)$, where $H = kf_{est}^{-1}$, to the desired linear transfer function $g(x)$ (see Figure 4 6) Usually and usefully the desired transfer function $g(x)$ is linear $g(x) = kx$ as in Figure 4 6

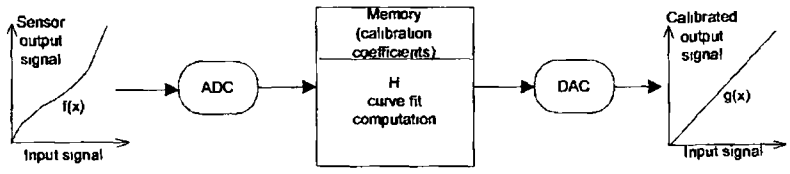

Figure 4 6 Inverse transfer function method for linearising calibration based on curve fitting

The expression for $h(x)$ depends on the specific curve fitting technique used Several curve fitting methods are further investigated in this section

Curve fitting can be achieved by either polynomial approximation or interpolation In function approximation the approximating function does not have to match the given data exactly Interpolation produces a function that matches the given data exactly at calibration points and also provides a good approximation to the unknown data values at intermediate points

In calibration more than 5 points will rarely be used so an interpolation would produce better results The only advantage is if noise is present (errors in measurement) In that case approximation might be better than matching the data exactly

Approximation methods are based on polynomials A polynomial of degree n is uniquely determined by $n+1$ points The number of points used determines the degree of the polynomial used The general form of a polynomial of degree n is

$$P(x) = c_n x^n + c_{n-1} x^{n-1} + \quad + c_1 x + c_0 \tag{4 1}$$

If the value of the function is given at several points it is possible to approximate that function The Weierstrass approximation theorem proves that [36]

Weierstrass approximation Theorem

Suppose that f is defined and continuous on $[a, b]$ For each $\varepsilon > 0$, there exists a polynomial $P(x)$ defined on $[a, b]$, with the property that

$$| f(x) - P(x) | < \varepsilon \tag{4 2}$$

for all $x \in [a, b]$

This theorem can be used in interpolation as well

Three approaches to interpolation are considered

- Taylor series interpolation,
- Solving the systems of equations using Gaussian elimination,
- Interpolation with Lagrange polynomials

4.4.1 Taylor Polynomial Interpolation

Taylor's Theorem

The function interpolation in one point a can be attained by the Taylor polynomial (equation 4 3) interpolation The Taylor polynomial interpolation matches the function in the single point a and trys to follow its derivatives at that point

Taylor's Theorem[37]
Let I be an interval containing the number a If f has n continuous derivatives on I, then

$$f(x) = f(a) + f'(a)(x - a) + \quad + \frac{f^{(n-1)}(a)}{(n-1)!}(x - a)^{n-1} + R_n(x) \tag{4 3}$$

where

$$R_n(x) = \frac{1}{(n-1)!} \int_a^x f^{(n)}(s)(x - s)^{n-1} ds \tag{4 4}$$

The remainder term $R_n(x)$ is the difference between a Taylor polynomial and a function it approximates, in other words, it defines the truncation error

Summary

Taylor polynomials can be used to approximate a function with a high degree of accuracy The main drawback of the method is that it requires equally spaced points due to utilising numerical methods for the derivatives computations as will be shown in the next chapter As a result, the reconstruction of a function on a large interval is sometimes not good enough simply by looking at its behaviour at a (one point) However, on small intervals the method is very accurate For more accuracy more points are needed

4.4 2 Gaussian Elimination Method

Finding the polynomial from equation 4 1 means finding the $n+1$ coefficients c_n, c_{n-1}, ,c_0 Since there are $n+1$ unknowns, $n+1$ equations are needed The equations are found simply by the fact that the polynomial $P(x_i) = y_i$ for $i = 0, 1, \quad , n$ This yields the system of equations

$$\begin{cases} c_n x_0^n + c_{n-1} x_0^{n-1} + \quad + c_1 x_0 + c_0 = y_0 \\ c_n x_1^n + c_{n-1} x_1^{n-1} + \quad + c_1 x_1 + c_0 = y_1 \\ \\ c_n x_n^n + c_{n-1} x_n^{n-1} + \quad + c_1 x_n + c_0 = y_n \end{cases}$$

which can be rewritten as a matrix equation

$$XC = Y \tag{4 5}$$

where

$$X = \begin{bmatrix} x_0^n & x_0^{n-1} & & x_0 & 1 \\ x_1^n & x_1^{n-1} & & x_1 & 1 \\ & & & & \\ x_n^n & x_n^{n-1} & & x_n & 1 \end{bmatrix}$$

is a data matrix

$$C = \begin{bmatrix} c_n \\ c_{n-1} \\ \\ c_0 \end{bmatrix}$$

is a vector of unknown coefficients to be determined and

$$Y = \begin{bmatrix} y_0 \\ y_1 \\ \\ y_n \end{bmatrix}$$

is a data vector

If the determinant $\det X \neq 0$, then the equation 4 5 has a unique solution

$$C = X^{-1}Y \tag{4 6}$$

X and Y are known matrices while C is the vector of coefficients

For example, interpolation with a third order polynomial of the form 4 1 is simply obtained solving the system of 4 equations with 4 unknown (4 coefficients) with basic Gaussian elimination The method consists of two phases upper triangulation (all entries of matrix X below its main diagonal are zero) or lower triangulation (all entries above its main diagonal are zero) and backward substitution Usually upper triangulation is chosen Often there is a need to rearrange the order of the equations during the upper triangulation of Gaussian elimination m order to reduce propagated round off error This is called pivoting and it is considered later on m Chapter 5

Summary

Though the method is quite simple, solving large systems is not efficient if the number of additions/subtractions, multiplications and divisions is taken into account [38] Therefore solving systems of equations should be avoided whenever possible

4 4.3 Interpolation with Lagrange Polynomials

Lagrange Polynomials

Lagrange interpolation is a method of finding a polynomial $y = f(x)$ which passes through a specified set of n points (x_i, y_i) such that $a \leq x_0 < x_1 < \quad < x_n \leq b$ and $y_i = f(x_i)$, for $i = 0, 1, \quad , n$ The Lagrange polynomial is defined as

$$L_{n,k}(x) = \prod_{i=0, i \neq k}^{n} \frac{x - x_i}{x_k - x_i} \tag{4 7}$$

The terms $L_{n,k}(x)$ (Lagrange polynomial coefficients) have the following properties

$$L_{n,k}(x_i) = \begin{cases} 1 & i = k \\ 0 & i \neq k \end{cases}$$

The Lagrange polynomial is the polynomial of form

$$P_n(x) = \sum_{k=0}^{n} L_{n,k}(x) y_k \tag{4 8}$$

with the degree n [39, 40]

Summary

This method solves the problem of large computations, making the calculations trivial The disadvantage of this technique is that the measurements have to be chosen with very small error, because there are no degrees of freedom left after interpolation, i e polynomial degree must be chosen in advance and a change of the degree involves recomputation of all terms For a polynomial of high degree the process involves a large number of multiplications and can become very slow

4.4.4 Error Analysis for Polynomial Interpolation

The basic equation for error analysis [41]

$$\textit{True value} = \textit{estimate} + \textit{error} \tag{4 9}$$

is simply applied when approximating a function f with the polynomial $P_n(x)$ The polynomial interpolation error is written as

$$R_n(x) = f(x) - P_n(x) \tag{4 10}$$

The error introduced by Taylor approximation is simply the remainder term presented in the equation 4 4 Rewriting the error term in the Taylor approximation [42], the error might be represented as follows

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1} \tag{4 11}$$

where c is the number between x and a

Additionally there is an approximation error in differentiation For example, the error of the second derivative would be

$$f''(x) = \frac{f(x-h) - 2f(x) + f(x+h)}{h^2} - \underbrace{\sum_{j=1}^{\infty} \frac{f^{(2j+2)}(x)}{(2j+2)!}h^{2j}}_{error} \tag{4 12}$$

The error associated with Lagrange polynomials is similar to the error term for Taylor polynomials and has the form [39]

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)(x-x_1) \quad (x-x_n) \tag{4 13}$$

where c is the value that lies in the interval $[a, b]$

4.4 5 Error Analysis Using Simulated Testing of Curve Fitting Method

Before going into detailed investigation of various curve fitting methods it was necessary to verify how good different curve fitting methods in function approximation are and what kind of results (errors) are to be expected In that manner, it was necessary to generate data points (from MATLAB) from known non-linear functions which have different non-linearities on the chosen interval and then test

different methods using MATLAB Simulations are performed on functions that change rapidly (extreme cases) $\sqrt{x}$, x^{-1}, $\log x$ and e^x and the results of the approximations are compared with original functions

Simulation results of interpolation methods in MATLAB are provided in Appendix B, C and D They showed that the advantage of the Gaussian method is the possibility to do the approximation between non equally spaced interpolation points and therefore obtain less error for logarithmic functions and x^{-1} type functions Much of the difficulty with polynomial interpolation can be avoided by interpolating not at equally spaced points but at the Chebyshev points [40] These points are mapped onto the interval $[a, b]$ in a way that places points more densely near the ends of the interval This is not possible to implement in Taylor approximation due to the restriction caused by numerical differentiation

Tests on the Taylor series interpolation method are performed on the second and third order Taylor polynomial Calibration coefficients for both polynomials are calculated for 5 points Accordingly, simulations for Gaussian elimination and Lagrange interpolation are carried out for the same number of points in order to be comparative

Lagrange interpolation and Taylor series gave better results than Gaussian elimination The error is large in a part of the range where a function changes rapidly However the reason for having good approximation results for the function e^x is that the simulation is carried out on the small interval $[0, 4]$ The function e^x changes rapidly When considered over a wide range the error is very large, approaching 10^4 at $x = 10$ and 10^{43} at $x = 100$ The small interval was chosen because a plot over the wide range is not legible

Taylor series interpolation of third order for 5 calibration measurements evaluated in the middle point gives good results only on the small interval around the approximating point whereas, for the second order Taylor polynomial evaluated in the middle point, the error on the full range is less but increases tremendously where the function changes rapidly On the rest of the range the approximation is very poor (compare the results in appendix) Increasing the number of calibration

Relationship	Transform	Substitution
$y = ax^b$	$ln(y) = \log(a) + b\log x$	$Y = \log(y)$ *and* $X = \log x$
$y = \frac{a}{bx + c}$	$\frac{1}{y} = \frac{b}{a}x + \frac{c}{a}$	$Y = \frac{1}{y}$ *and* $X = x$
$y = \frac{1}{(ax + b)^2}$	$\frac{1}{\sqrt{y}} = ax + b$	$Y = \frac{1}{\sqrt{y}}$ *and* $X = x$

Table 4 1 Function transformation

measurements results in reducing the error Lagrange interpolation on the other hand has a tandancy to oscillate for higher order polynomials

One possibility to overcome polynomial interpolation problems is function transformation - sensor characteristic linearisation In this case some knowledge of the sensor response is necessary (see Table 4 1) The function is first transformed into a linear function allowing easier calculation of the linearisation coefficients Once the coefficients are calculated it is possible to determine the inverse transfer function algebraically

The most important step in an approximation problem is to choose a good model (appropriate method) for the data

4.5 Linearisation and Calibration Based on Progressive Polynomial Calibration Method

The progresive polynomial calibration (PPC) and linearisation method operates on the principle that each calibration measurement can be used directly to cal-

culate one calibration coefficient in the correction function [27] The correction basically modifies the sensor output The next calibration step is based on this correction in a way that each succeeding correction is independent of the current one

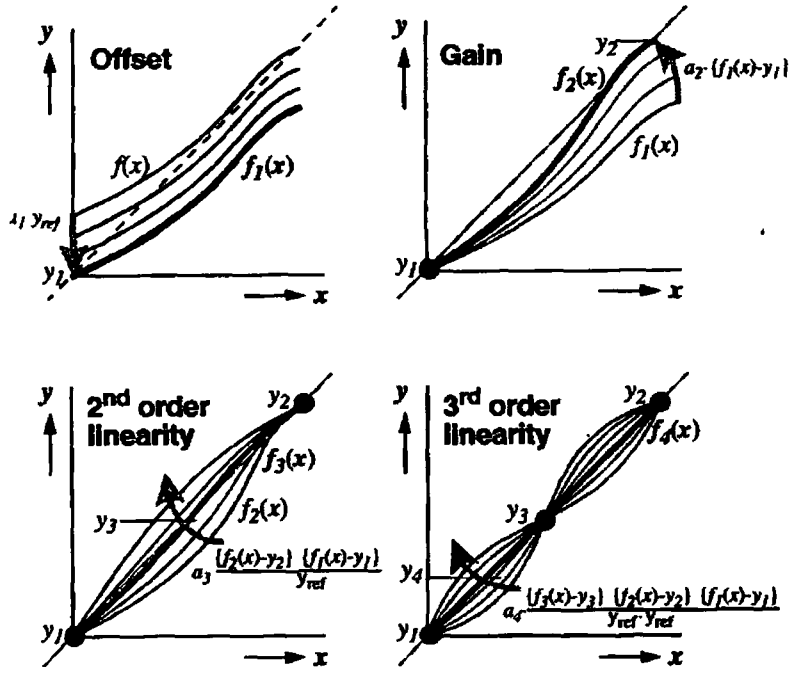

Figure 4 7 Four steps of the progressive polynomial calibration method

The first measurement is used to eliminate the offset This means the transfer function is shifted or translated (Figure 4 7 [27]) The second measurement corrects the gain error without affecting the offset calibration It rotates the function around the previous calibration point Further measurements are used to correct linearity which is achieved by bending the function while the previous calibration points stay fixed Thus the polynomial curve progresses towards the desired transfer function (see Figure 4 7), hence the name progressive polynomial calibration (PPC)

4 5.1 Mathematical Approach to PPC

The input variable to the sensor is indicated as x, the sensor output is indicated by y, the sensor transfer function is denoted as $y = f(x)$, the desired sensor transfer function is $y = g(x)$ and it is assumed to be linear function of the input signal, $g(x) = Kx$ (Figure 4 8)

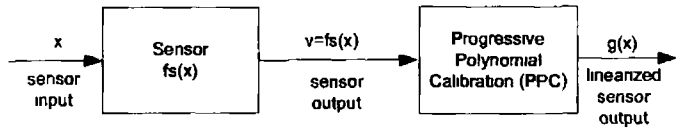

Figure 4 8 Block diagram of PPC method

PPC is performed in order to get a calibrated sensor response function that closely resembles the desired transfer function This is done by taking calibration measurements for a set of a well-known input signals x_n, and comparing the measured sensor output $f(x_n)$ to the desired output The method [27] calculates a corrected transfer curve $h_n(x)$ after each calibration measurement Previously calculated transfer functions $h_1(r)$ to $h_{n-1}(r)$, the previous calibrated output values y_1 to y_n and the n^{th} calibration measurement $f(x_n)$ are used in calculations The calibration can be presented as a series of nested formulae

$$h_n(a_n, h_1(x), \quad , h_{n-1}(x), y_1, \quad y_{n-1}) \tag{4 14}$$

where a_n is referred to the *nth* calibration coefficient At each calibration step the calibration coefficient a_n has to be calculated in order to obtain the linearised sensor output $h_n(x_n) = g(x_n) = y_n$

The procedure for calculating the first four steps is as follows

1 A reference signal y_{ref} is needed only to give the correction factor, so that calibration coefficients a_n are nondimensional numbers Based on the first calibration measurement $f(x_1)$ and the desired output value $y_1 = g(x_1)$, the first calibration coefficient a_1 and the first calibration function $h_1(x)$ can be calculated

$$a_1 = \frac{y_1 - f(x_1)}{y_{ref}} \tag{4 15}$$

$$h_1(x) = f(x) + a_1 y_{ref} \tag{4 16}$$

Evaluating h_1 for x_1 (equation 4 16) shows that $h_1(x_1) = y_1$ and the desired correction is obtained (offset elimination)

2 The second calibration coefficient uses the previously corrected transfer function h_1 in points x_1 and x_2

$$a_2 = \frac{y_2 - h_1(x_2)}{h_1(x_2) - y_1} \tag{4 17}$$

The second calibration coefficient a_2 and previously calculated corrected transfer function $h_1(x)$ are used for calculating the second corrected transfer function

$$h_2(x) = h_1(x) + a_2\{h_1(x) - y_1\} \tag{4 18}$$

Again letting $x = x_2$ in equation 4 18 will result in $h_2(x_2) = y_2$ Also, since $h_1(x_1) = y_1$ equation 4 18 becomes $h_2(x_1) = h_1(x_1) = y_1$ which means total independence of a_2 for the first calibration measurement leaving the previous calibration intact (gain elimination)

3 The third calibration function gives a second order polynomial correction

Similarly $h_1(x_3)$ and $h_2(x_3)$ are used to find the third calculation coefficient

$$a_3 = \frac{\{y_3 - h_2(x_3)\}y_{ref}}{\{h_1(x_3) - y_1\}\{h_2(x_3) - y_2\}} \tag{4 19}$$

This results in new corrected transfer curve $h_3(x)$

$$h_3(x) = h_2(x) + a_3\frac{\{h_1(x) - y_1\}\{h_2(x) - y_2\}}{y_{ref}} \tag{4 20}$$

Thus it is possible to show that the corrected transfer function in previously calibrated points is independent of a_3 (non-linearity correction)

4 Analogously, a_4 and $h_4(x)$ (non-linearity correction) are

$$a_4 = \frac{\{y_4 - h_3(x_4)\}y_{ref}^2}{\{h_1(x_4) - y_1\}\{h_2(x_4) - y_2\}\{h_3(x_4) - y_3\}} \tag{4 21}$$

and

$$h_4(x) = h_3(x) + a_4 \frac{\{h_1(x) - y_1\}\{h_2(x) - y_2\}\{h_3(x) - y_3\}}{y_{ref}^2} \qquad (4\ 22)$$

The equations can be generalized The calibration coefficient a_n is calculated as

$$a_n = \frac{y_n - h_{n-1}(x_n)}{y_{ref}} \prod_{i=1}^{n-1} \left(\frac{y_{ref}}{h_i(x_n) - y_i} \right) \qquad (4\ 23)$$

while the *nth* calibration function is given by

$$h_n(x) = h_{n-1}(x) + a_n y_{ref} \prod_{i=1}^{n-1} \left(\frac{h_i(x) - y_i}{y_{ref}} \right) \qquad (4\ 24)$$

4.5.2 Error Analysis Using Simulated Testing of PPC

The error is simply defined as the difference between the calibrated output and the desired output

$$\varepsilon(x) = h_n(x) - g(x) \qquad (4\ 25)$$

It depends on the number of steps and the choice of the calibration points

1. When taking only one calibration step, the best results are accomplished when the calibration measurement r_1 is taken at the middle of the input range

2. When taking two calibration steps (two calibration coefficients), the calibration measurements should be taken at about 25% and 75% of the input range

3. Taking three or more calibration measurements gives a good linearisation The calibration points should be taken in the following order the first point at one end of the sensor range, the second point at the other end of the sensor range, and further calibration points half way between two previously selected points

For example, if the sensor transfer function is $f(x) = \ln x$ and the calibration measurements are $x_1 = 1$, $x_2 = 9$, $x_3 = 5$, $x_4 = 7$ and $x_5 = 3$ then it is possible to

do the error analysis and see how the number of measurements affects the error Suppose the desired sensor transfer function is $y = g(x) = 0\ 25x - 0\ 25$ (chosen in such a way so the coefficient $a_1 = 0$ to simplify the calculation), which gives the desired values $y_1 = 0$, $y_2 = 2$, $y_3 = 1$, $y_4 = 1\ 5$ and $y_5 = 0\ 5$ (Figure 4 9)

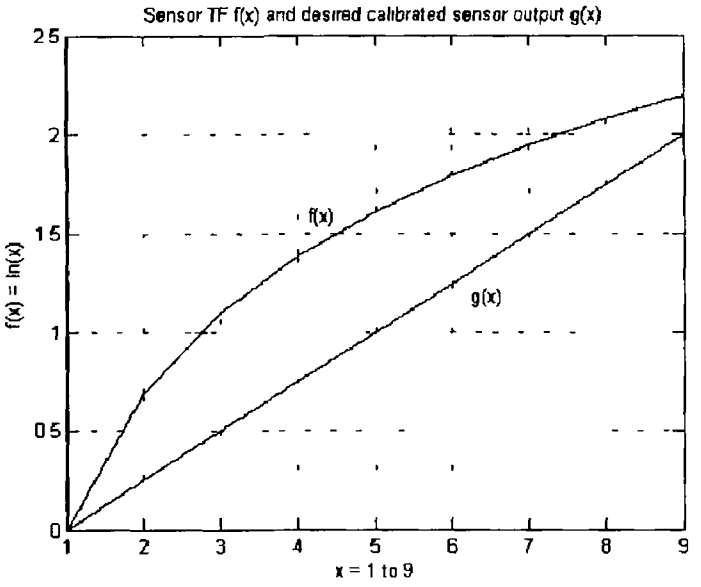

Figure 4 9 Sensor transfer function $f(x)$ and desired sensor transfer function $g(x)$

The first calibration coefficient is $a_1 = 0$ as the sensor output at x_1 matches its desired transfer curve $f(x_1) = g(x_1)$ The second calibration coefficient is used to adjust the gain of the sensor output to the desired calibration curve (Figure 4 10) whereas the rest of the coefficients have an effect on the linearity

The actual error of each calibration step is shown in Figure 4 11 Increasing the number of calibration coefficients reduces the linearity error

The error is the largest about point 1 5 (Figure 4 11) and it is about 0 73% By not adhering to rule 3, it is possible to reduce the error (i e at the point $x = 1\ 5$) about that point bringing the increase on the overall range (Figure 4 12)

The results of the simulation suggest taking the next calibration point at the point $x = 1\ 5$ where the error is the greatest, thus reducing the error to less then

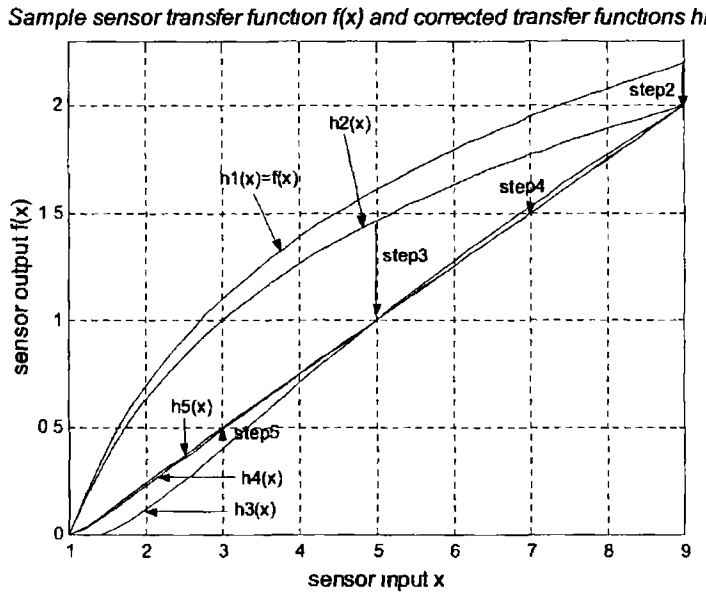

Figure 4 10 Sample sensor transfer function $f(x)$ and corrected transfer functions $h_n(x)$

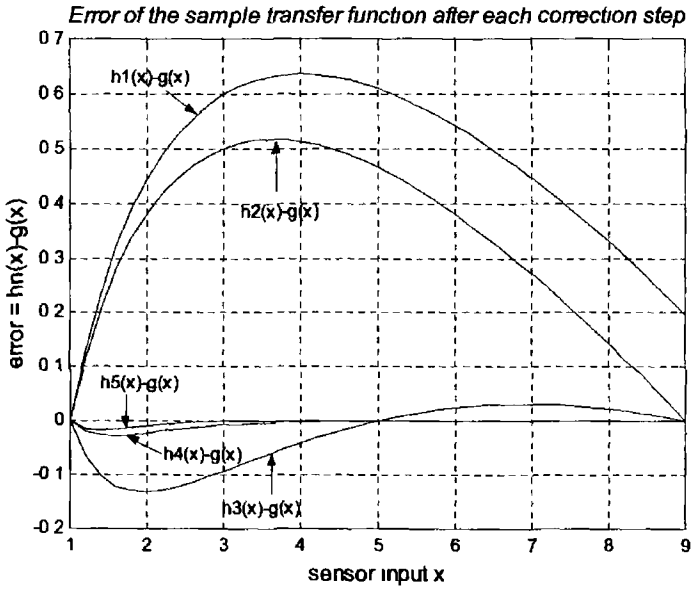

Figure 4 11 Error of the sample transfer function, after 1, 2, 3, 4 and 5 correction step

0 2% (Figure 4 13) The smoothness of the curves can be improved by selecting a smaller step in MATLAB

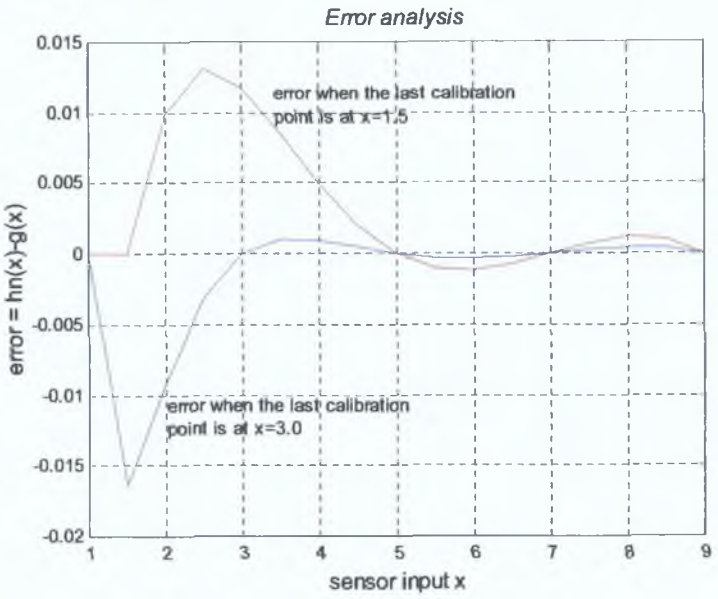

Figure 4.12: Error analysis.

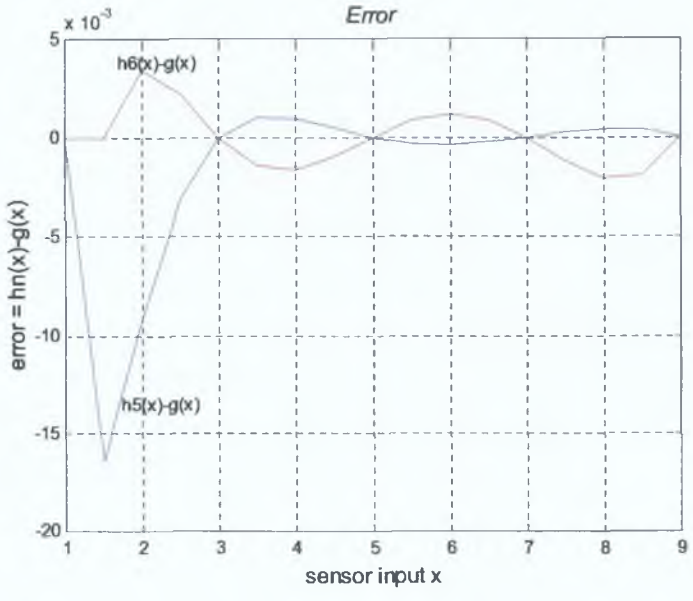

Figure 4.13: Error analysis when one more calibration measurement is added.

The linearity that is possible to achieve depends on the non-linearity of the sensor transfer function. If the sensor transfer function is known, calibration points can be chosen based on the aforementioned rules and/or experience, hence yielding better results. The calibration measurements can be then optimized accordingly

for better results

The choice of the desired linearised curve is arbitrary and does not have any effect on an error Yet, it makes sense to make a decision on the desired values beforehand These values would be stored in microcontroller and would be universal for any sensor transfer function Similarly, the reference signal y_{ref} can be treated the same way

Chapter 5

Modelling and Simulation of Linearisation Methods

In this chapter, linearisation methods using function interpolation based on Taylor polynomial interpolation, Gaussian elimination, Lagrange interpolation and progressive polynomial calibration (PPC) are presented as well as some models and simulations These methods were chosen for interpolation since polynomials are a good deal simpler than most functions we encounter They are easier to evaluate at particular values and easier to differentiate and integrate

It is possible to use curve fitting to get a good approximation of a function from a set of data (x_i, v_i), $i = 0, 1, \quad , n$ Once an approximation is performed it is possible to determine an inverse function analytically Since the sensor transfer function is more often than not unknown, it is better to use curve fitting directly on the set of the data (v_i, x_i) to get a good approximation of $v = f^{-1}(x)$ The Taylor polynomial interpolation, Gaussian elimination and Lagrange interpolation were implemented by going through each (v, x) point (Figure 4 1) and approximating the inverse transfer function of the sensor transfer function (§4 4) Progressive polynomial calibration was simulated in SIMULINK, some unexpected difficulties were encountered and these are discussed in this section The goal was the implementation of these algorithms in C code and their implementation on a Microchip 8-bit PIC16F877 microcontroller i e from theory to actual application (sensor system)

5.1 Taylor Series Implementation

Every analytic function can be represented by a power series such as a Taylor series [43] Probably the most common application of the Taylor series is to use the Taylor polynomial (the truncated Taylor series) as an approximation to the function It is very common in some areas to use a first (which is equivalent to approximating the function by the tangent line at the point) or second degree Taylor polynomial as the approximation However, since a number of sensors' characteristics are best approximated with the third order polynomial it was decided to implement the third order Taylor polynomial Thus for sensors whose outputs are better approximated by the second order polynomials the polynomial coefficient of the cubed term will be very small or zero

The advantages of the Taylor series approach are clear The truncation error (Equation 4 4) can be controlled by evaluating the derivatives to appropriate order This is always possible but frequently grows cumbersome for the higher derivatives, since the successive derivatives for many functions become progressively more complicated

To determine the coefficients of the third order Taylor polynomial, i e , the first three derivatives, the numerical finite difference formulae were applied Estimation of derivatives at some point (x_k, y_k) can be made to the left (forward differences), to the right (backward differences) or about the middle point (central differences) Finite differences called central differences use the slope of the secant through the points (x_{k-1}, y_{k-1}) and (x_{k+1}, y_{k+1}) instead of the slope of the tangent at the point (x_k, y_k), because it is generally more accurate, due to the symmetrical position of two points about the point of interest (Figure 5 1)

The disadvantage of the central differences is that they cannot be applied at the end points because there are no "known" points from both sides Estimating the *mth* derivative (m = 1,2 and 3 - for a third degree Taylor polynomial) at the middle point with equally spaced function values is considered The following are

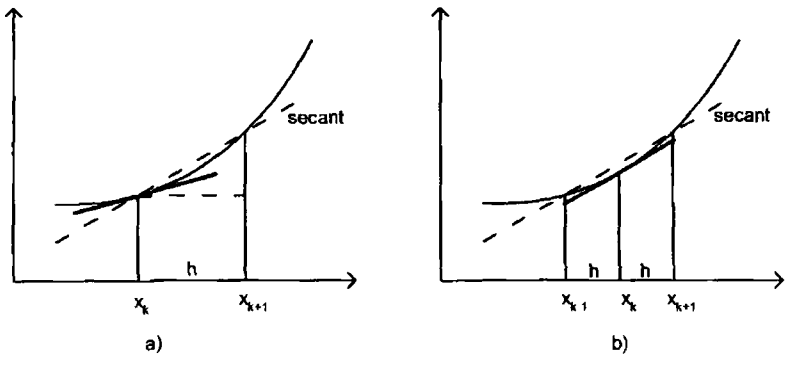

Figure 5 1 a) Backward differences, b) Central differences

the central differences formulae [41]

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h} \tag{5 1}$$

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2} \tag{5 2}$$

$$f'''(r) \approx \frac{f(x+2h) - 2f(x+h) + 2f(x-h) - f(x-2h)}{2h^3} \tag{5 3}$$

There are two sources of error in the derivative formulae above truncation error and round off error The truncation error comes from higher terms in the Taylor series expansion (thus it is clear that for $f'(x)$, for example, the truncation error comes from terms h^{2j} and higher as in the equation 5 6 below) whilst round off error has various contributions (for example, the numbers are rarely exactly represented in binary and they are represented with some fractional error of the machine's floating-point arithmetic)

Suppose we want the derivative of the function $f(x)$ at the point x and that the points $x+h$ and $x-h$ lie inside an interval $[a,b]$ [44] Then, using Taylor series, we can write

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!}f''(x) + \frac{h^3}{3!}f'''(x) + \tag{5 4}$$

and

$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2!}f''(x) - \frac{h^3}{3!}f'''(x) + \qquad (5\ 5)$$

subtracting (5 5) from (5 4) and dividing by $2h$ results in the equation

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} - \sum_{j=1}^{\infty} \frac{f^{(2j+1)}(x)}{(2j+1)!}h^{2j} \qquad (5\ 6)$$

Ideally $h \rightarrow 0$ However, if the calculation is being performed using a microcontroller, for example, then round off error must be considered In such a case numbers are represented with some fractional error that is characteristic of the microcontroller's floating-point format and if h is too small the round off error can dominate the approximation Obviously, having h too large results in large errors because of the inadequate model

Consistent with above, h should be chosen as small as possible in order to minimize the error that is inherent in all numerical methods

These derivatives are coefficients of the Taylor series and are used in further calculations for any other values of output from the sensor In order to make this method work it is necessary to have data points equally spaced ($x - nh$, $x + nh$ in equations 5 1- 5 2) for the calibration process

5.1.1 MATLAB and C Implementation of Taylor Series, Test and Validation

In attempting to set a benchmark against which the Taylor curve-fitting approximation could be compared the MATLAB 'taylor' built-in function from the Symbolic Toolbox was investigated The function is given as

taylor(f,n,x,a)

where f is the investigated function, x is the independent variable, n gives $n-1$ order polynomial approximation of f and a is the point at which the approximation was made

The idea was to compare the MATLAB generated coefficients with the corresponding coefficients obtained from, say, a C implementation of the approximating polynomial

The Taylor series was implemented in MATLAB using the Symbolic Toolbox First, a symbolic variable x and a function investigated are defined, and then the 'taylor' function is applied and evaluated for the middle point The example is shown in Figure 5 2 presenting the function $\log x$ that was interpolated with the third order Taylor polynomial and evaluated at the middle point ($x = 50$)

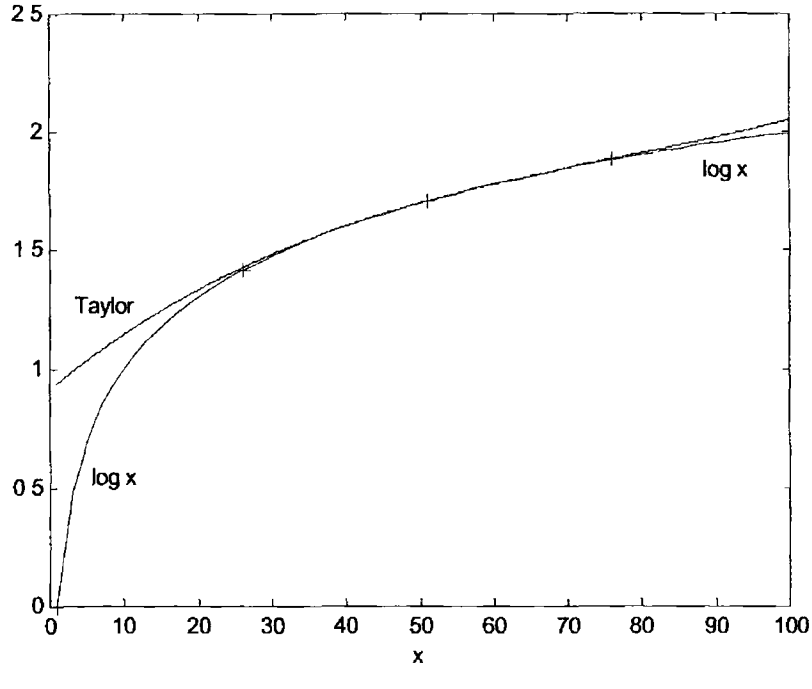

Figure 5 2 Third order Taylor series about point $x = 50$ for the function $\log x$ in MATLAB

The same procedure was followed for a number of different functions

Then the Taylor series technique was implemented in the C code on a PC (Personal Computer) The inverse function is required for linearisation of the output sensor characteristics (see Chapter 4) The required inversion is achieved by traversing the two axes of the function i e the axes are symmetrically mapped (like a mirror image) across the $y = x$ line In terms of C code this simply means swapping x and y in the output sensor characteristic and proceeding with the

Taylor series approximation of x as a function of y Implementation in C on the PC illustrated the limitations associated with this method One of the limitations is that a first and a second derivative (which are calibration coefficients) cannot be calculated at the end values of the array (the first and the last value of the sensor output) i e formulae (5 1 - 5 3) require output function values in $x - h$ and $x + h$ and similarly for the third derivative at $x - 2h$, $x - h$, $x + h$ and $x + 2h$

The C code is tested for 5 calibration measurements This means that only in the middle point (middle of the range) is it possible to calculate the third derivative and therefore obtain the third order Taylor polynomial A second order Taylor series in Figure 5 3 is implemented in 5 points as well as the third order in order to be able to compare these results Evaluation is performed with reference to the middle point The derivatives are calculated for the equal step h

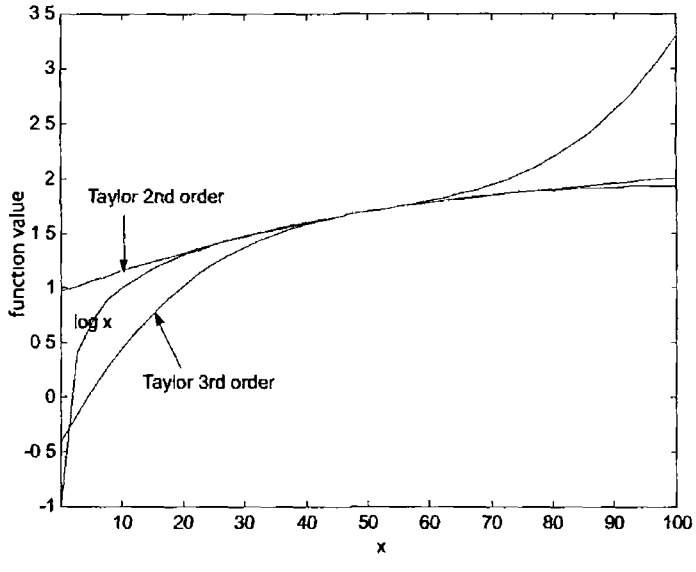

Figure 5 3 Third and second order Taylor series with one reference point for the function $\log x$ implemented on PC in C

It was also discovered that results from the MATLAB simulations and results from C differ greatly (Figure 5 2 and Figure 5 3 respectively)

The reason is that derivatives calculations in C are based on numerical differences (equations 5 1 - 5 3) while MATLAB has a built in Taylor function that uses a

symbolic toolbox This basically means that MATLAB does not do any approximations when calculating derivatives On the other hand programming languages can not avail of symbolic toolboxes (computer algorithms are based on numerical methods) The error, as it was described earlier, originates in choosing a step h, for 5 calibration measurements on a wide range, a large step h is inevitable Thus the approximation is only good around the approximating point while for the rest of the range it is very poor

The calibration coefficients (derivatives at the middle point) for a range of functions are calculated in MATLAB and C, and presented in a Table 5 1 and Table 5 2

MATLAB	f'	f''	f'''
$\sqrt{x}$	0 070711	0 000354	0 000003
x^{-1}	0 038447	0 007538	0 001478
$\log x$	0 008668	0 000086	0 000001
e^x	7 389056	3 694528	1 231509

Table 5 1 Calibration coefficients calculated in MATLAB

C	f'	f''	f'''
$\sqrt{x}$	0 073206	0 000385	0 000014
x^{-1}	0 050600	0 019840	0 300640
$\log x$	0 009518	0 000099	0 000011
e^x	8 683600	4 012800	1 571983

Table 5 2 Calibration coefficients calculated in C

It is evident that the error is large and it is not repetitive by any template Hence it is not predictable and cannot be compensated in any way by software

Since the approximation at one point in C did not show satisfying results, it was decided to perform curve fitting using the piecewise Taylor polynomial interpolation The measurements are taken at $x = 1, 2, \quad , 10$ and estimation is made for 6

middle points (6 Taylor polynomials) where it was possible (due to the numerical equations, i e $x - 2h$, $x = 2h$) to calculate the third order Taylor polynomials Figure 5 4 shows only points for which the estimation was made

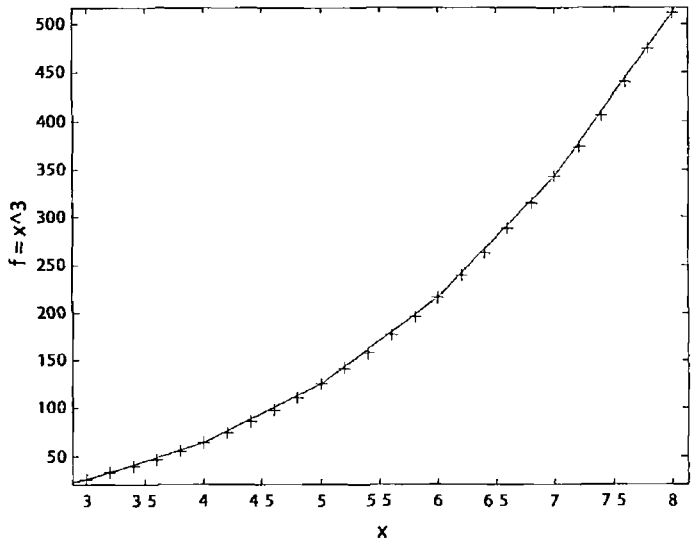

Figure 5 4 Estimated function values in C using the piecewise Taylor polynomials in 6 points

The compiled C code is tested for several different functions in order to validate the accuracy of the algorithm An error between theoretical and estimated function values using the piecewise Taylor series from Figure 5 4 are shown in Figure 5 5

The simulation showed that the error of the piecewise Taylor series approximation is acceptable and that it is linear on the chosen interval (Figure 5 5) This is done for 10 calibration points that enables storing coefficients of Taylor polynomials of third order for 6 calibration measurements (two measurements on both ends are not taken into account because of the limitation on formulae) For the points 3-8 inclusive, there are no errors found (as expected) because Taylor series coefficients are initially calculated from these measured values

Comparing all the results the most effective algorithm for the piecewise Taylor polynomial interpolation is derived If calibration measurements x are taken

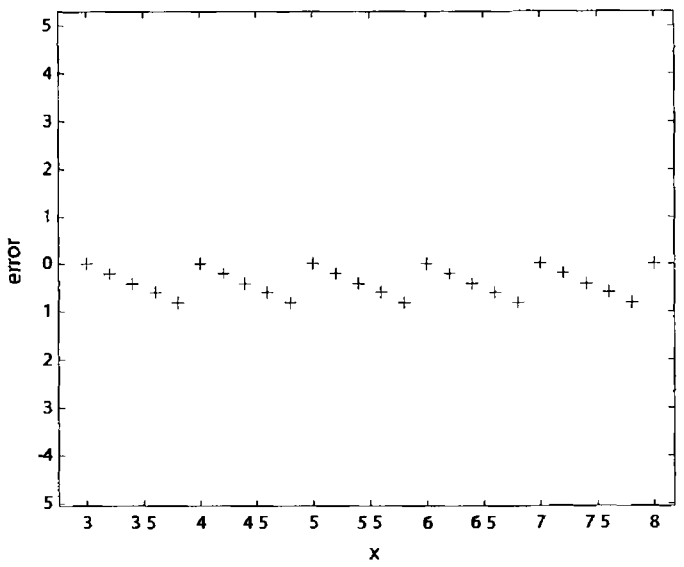

Figure 5 5 Error between theoretical and estimated function values using the piecewise Taylor series approximation

on the interval $x \in [a, b]$ and $a = x_1 < x_2 < x_3 < x_4 < x_5 = b$ (where x_n are calibration measurements) then evaluation of the Taylor polynomial should be performed in such a way that x_2 is used as a reference point for the values $x \in [x_1, x_3)$ (or represented as two subintervals $x \in [x_1, x_2)$ and $x \in [x_2, x_3)$), x_3 for the values $x \in [x_3, x_4)$ and x_4 for the values $x \in [x_4, x_5]$ This basically means that when evaluating a Taylor polynomial at the point a in the equation 4 3, the reference pomt a is chosen as

$$a = \begin{cases} x_{i+1} & \text{for } i = 1 \\ x_i & \text{for } i = 2, \quad ,n \end{cases} \tag{5 7}$$

where n is number of calibration measurements

Up to this point all "data" was generated from known mathematical functions As a further test it was decided to produce a set of data from a physical system that is known to have a non-linear response The system consists of the diode circuit (see Figure 5 6) whose input voltage can be varied from 1-5V and the response of the circuit recorded These input values and corresponding output values are used as input data for calculating the Taylor polynomial approximation to the voltage

transfer characteristic (VTC) of this circuit using the C code implementation on the PC The diode used is the standard 1N914 diode The test once again validated the method derived by the equation 5 7

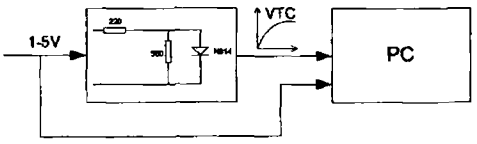

Figure 5 6 Test system

However, ten is a lot of measurements and calculating piecewise Taylor series is not very practical It also requires a lot of memory to store the coefficients of each subrange

5 1 2 Implementation and test on PIC

With the C code of the Taylor series on the PC yielding good results it was then necessary to implement the C code on a embedded system The Microchip PIC16F877 microcontroller [45] was used for this purpose to perform the Taylor polynomial interpolation It uses a platform called MPLAB to write, compile and program the devices

The experiment was set up as shown in Figure 5 7, using the diode test circuit again

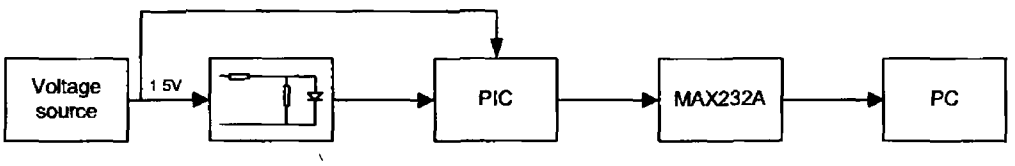

Figure 5 7 Taylor series test circuit for PIC

The PIC16F877 microcontroller (Figure 5 8) has a 10-bit ADC (analogue to digital converter) The input voltage of the diode circuit is connected to one of the

analogue channels of the PIC where it is transformed into digital form allowing the calculation of the inverse transfer function coefficients The PIC timer is configured as a 1s counter When 10s is counted (time allowed for changing the input voltage of the circuit), the interrupt routine is called, the ADC reads the value and stores it in an array Set output voltages are pre-written in the C code in the form of an array as well (an algorithm applied calculates the Taylor polynomial of the inverse VTC function) These values are then used in the calculations

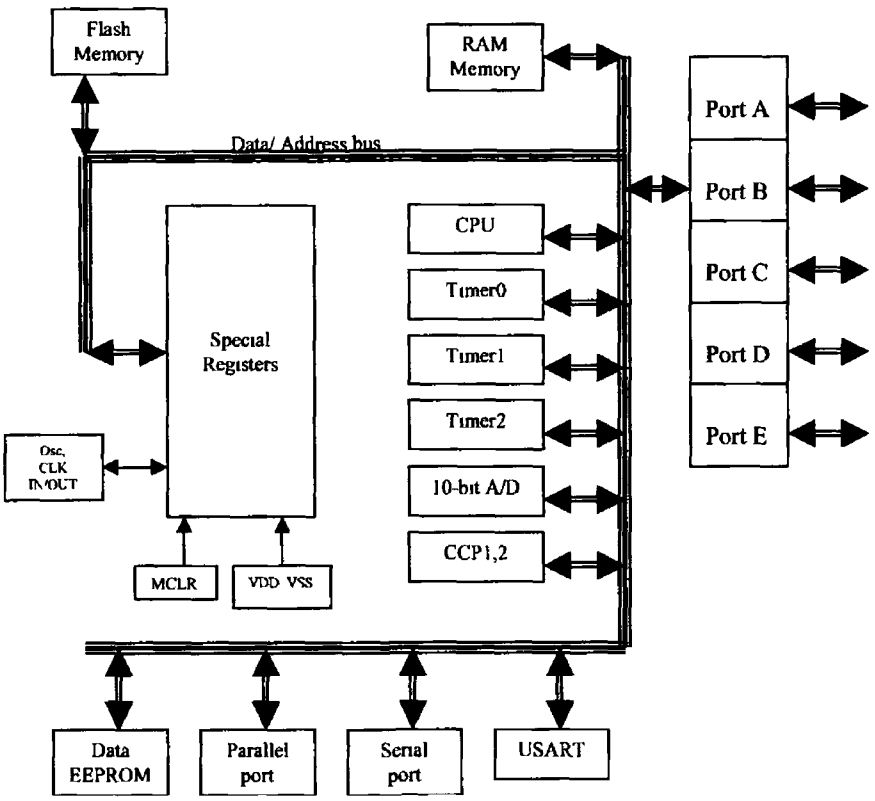

Figure 5 8 Simple block diagram of PIC16F877

To check the results obtained it was necessary to find a way of displaying them It was decided to establish RS232 communication (mentioned earlier in Chapter 2) between the PIC and the PC where the results would be displayed The PIC16F877 has a built-in serial communication interface, a USART (Universal Synchronous Receiver Transmitter) module, which needed to be connected to MAXIM MAX232A chip [46] (Multichannel RS232 Drivers/Receivers) before connecting it to the RS232 cable of the PC This is required to cope with voltage level differences between the PIC's serial communication interface and the PC's

RS232 serial communication port

The reason why this type of the communication is chosen is because the communication between PIC and PC is also necessary during a calibration process (first time calibration) to store the input sensor values onto the PIC while measuring the sensor output Later, for recalibration, that connection is not needed and it will work as a stand-alone device, which is the end goal

Calculated voltage input	**Set value of the output voltage of the diode circuit**	**Measured voltage input**	**Error**
0 418V	0 3V	0 419V	0 001V
0 453V	0 325V		
0 488V	0 35V	0 489V	0 001V
0 524V	0 375V		
0 560V	0 4V	0 561V	0 001V
0 596V	0 425V		
0 632V	0 45V	0 633V	0 001V
0 673V	0 475V		
0 716V	0 5V	0 717V	0 001V
0 767V	0 525V		
0 823V	0 55V	0 823V	0 000V

Table 5 3 Results of the experiment

The C code was successfully compiled and loaded onto the PIC for testing The first step was to read a value of the signal, perform ADC conversion, store it into the array and send it by serial transmission pin on the PIC to PC via RS232 communication using the MAX232A Ten measurements all together were taken and used for calculating the Taylor series coefficients of the inverse function on the PIC After the coefficients were calculated they were sent via the serial link to the PC The values displayed on the PC were then validated The results are summarised in the Table 5 3 Measured voltage input is the actual input of the diode circuit and calculated voltage input is the result of Taylor series calculation (estimated value using Taylor polynomial coefficients) at set values of the diode

circuit output voltage and at points in between the "known" points The error introduced is due to machine precision (round off error)

5.2 Gaussian Elimination Implementation

Gaussian elimination is applied on the inverse function (axis swapping) for linearisation purposes Finding the inverse matrix is an inefficient approach in computer algorithms Basic elimination without pivoting consists of $n-1$ eliminations steps for a system of n equations and gives the transformed linear system (equation 5 13) equivalent to the original system (equation 4 5) that is then solved by backward substitution The strategy is to reduce one column at the time to form an equivalent upper-triangular system (equation 5 13)

"Data" matrix X in equation 4 5 can be written as an extended (known) or augmented matrix $[X|Y]$ as

$$[X|Y] = \begin{bmatrix} x_{11} & x_{12} & & x_{1n} & y_1 \\ x_{21} & x_{22} & & x_{2n} & y_2 \\ & & & & \\ x_{n1} & x_{n2} & & x_{nn} & y_n \end{bmatrix} \tag{5 8}$$

At each elimination step new elements of the $[X|Y]$ matrix are calculated and old ones are overwriten, i e the k^{th} step reduces the k^{th} column of X by subtracting a multiple of the k^{th} row from the subsequent rows to give zeros below the k^{th} diagonal element $x_{ik}(old)$ is replaced by $x_{ik}(new)$ as follows

$$x_{ik}(new) = \frac{x_{ik}(old)}{x_{kk}} \tag{5 9}$$

A "shorthand" method of describing this process is $=$, so it is possible to write

for $i = k+1, \quad , n$

$$x_{ik} = \frac{x_{ik}}{x_{kk}} \tag{5 10}$$

for $j = k+1, \quad , n$

$$x_{ij} = x_{ij} - x_{ik}x_{kj} \tag{5 11}$$

$$y_i = y_i - x_{ik}y_k \tag{5 12}$$

After upper triangulation the extended matrix $[X|Y]$ will have become $[X_{new}|Y_{new}]$

$$[X_{new}|Y_{new}] = \begin{bmatrix} x_{11} & x_{12} & & x_{1n} & y_1 \\ 0 & x_{21} & & x_{2n} & y_2 \\ & & & & \\ 0 & 0 & & x_{nn} & y_n \end{bmatrix} \tag{5 13}$$

Equation 4 5 then becomes

$$UC = W \tag{5 14}$$

where $U = X_{new}$ is the upper triangular matrix, with overwritten values X, and $W = Y_{new}$ (both known) and C is vector of coefficients

This system is solved by backward substitution

$$c_n = \frac{y_n}{x_{nn}} \tag{5 15}$$

for $i = n-1, n-2, \quad , 1$

$$c_i = (y_i - \sum_{k=i+1}^{n} x_{ik}c_k)/x_{ii} \tag{5 16}$$

The computational approach is to use upper triangulation and backward substitution or lower triangulation and forward substitution instead of solving the matrix equation given by 4 6

5.2.1 Pivoting Strategies

The element x_{kk}, by which all subsequent column elements have to be divided within k^{th} step, is called the pivot Applying the above described algorithm it is possible to find the necessary multiplier to reduce each column to zero If the pivot element is zero, depending on whether or not there are any zero elements in the pivot column below the pivot row, two situations may arise If the entire pivot column below the pivot row is zero, the system of equations does not have a unique solution In the other case when there is a nonzero element in the pivot column below the pivot element, the algorithm for Gaussian elimination can be modified to avoid dividing by zero This can be simply achieved by pivoting, i e interchanging rows so that one has a nonzero diagonal element [41] However, in

practice this is not sufficient and not only zero, but also small pivots have to be avoided whenever it is possible Therefore when interchanging rows the element with the largest absolute value on or below the k^{th} diagonal is taken

$$|x_{kk}| = max|x_{ik}| \quad \text{for } i = k, k+1, \quad , n \tag{5 17}$$

This is called row or partial pivoting For example, the system given by

$$\begin{aligned} 0\,0001x + y &= 1 \\ x + y &= 2 \end{aligned}$$

has a solution $x = 1\,0001$ and $y = 0\,9999$ using exact arithmetic If pivoting is not applied then, using three-digit arithmetic, the system becomes

$$\begin{aligned} 0\,0001x + y &= 1 \\ -10000y &= -10000 \end{aligned}$$

with the solutions $x = 0$ and $y = 1$ The result is rounded and the solution is quite a good approximation of the first equation while it is a very bad approximation of the second equation Pivoting the rows with the largest absolute value on or below the k^{th} diagonal, system becomes

$$\begin{aligned} x + y &= 2 \\ y &= 1 \end{aligned}$$

having solutions $x = 1$ and $y = 1$ which is good enough for using three-digit arithmetic

If the largest absolute value is zero then the matrix is singular and the elimination is stopped, or if it is smaller than an accuracy tolerance ε (equation 5 18), then the matrix is numerically singular and again the elimination procedure is stopped For that reason, the size of pivots determines the stability of the algorithm as far as the development of roundoff errors is concerned

$$|x_{kk}| < \varepsilon \tag{5 18}$$

The other pivoting technique is complete or full pivoting In this case both rows and columns are interchanged so the largest number is used as pivot

$$|x_{kk}| = max|x_{ij}| \quad \text{for } i,j = k, k+1, \quad ,n \tag{5 19}$$

5.2.2 The Full Gaussian Algorithm

The Gaussian elimination can be presented as the following steps

1 Selecting the pivot step using rule given by the equation 5 17,

2 Row pivoting and elimination,

3 Backward substitution

The algorithm was first implemented in MATLAB for a simulation and then the C code was written and tested on the PC and PIC The results of the simulation are in Appendix C

The accuracy of the algorithm may be further improved by iterative refinement [47] The algorithm implemented and used for this project uses the maximum element as a pivot element (equation 5 17) and does not include the iterative refinement

5.3 Lagrange Interpolation Implementation

Linearisation using Lagrange interpolation can be carried out the same way as the Taylor polynomial interpolation, by traversing the axes of the sensor transfer function

In order to implement Lagrange interpolation in MATLAB and similarly in C and make it easier for programming, equations 4 7 and 4 8 were rearranged $L_{n,k}(x)$ can be written as

$$L_{n,k}(x) = \frac{(x - x_1) \quad (x - x_{k-1})(x - x_{k+1}) \quad (x - x_n)}{(x_k - x_1) \quad (x_k - x_{k-1})(x_k - x_{k+1}) \quad (x - x_n)} \tag{5 20}$$

The numerator is the product

$$N_k(x) = (x - x_1) \quad (x - x_{k-1})(x - x_{k+1}) \quad (x - x_n) \tag{5 21}$$

and the denominator is of the same form, but with the variable x replaced by the given value x_k

$$D_k(x) = (x_k - x_1) \quad (x_k - x_{k-1})(x_k - x_{k+1}) \quad (x - x_n) \tag{5 22}$$

then the Lagrange interpolation polynomial can be written in the form

$$p(x) = c_1 N_1 + c_2 N_2 + \quad + c_n N_n \tag{5 23}$$

where

$$c_k = \frac{y_k}{D_k} = \frac{y_k}{(x_k - x_1) \quad (x_k - x_{k-1})(x_k - x_{k+1}) \quad (x - x_n)} \tag{5 24}$$

and N_k is already given by the equation 5 21

Equations 5 24 and 5 21 are implemented in both MATLAB and C The numerical algorithm in MATLAB is used rather then the Symbolic Toolbox

In general the results showed that equally spaced y_k points (sensor output) give good results when interpolating the inverse sensor transfer function The proposed number of measurements is 5 and more than 6 is not advisable as it causes large numerical "oscillations", like for the Runges function (Figure 5 9) For the function x^{-1} an additional sixth measurement in the region of rapid changes will give better results The simulation results obtained can be found in the Appendix D

The question imposed was what degree polynomial should be used for the approximation One would expect that the approximating polynomial will get very close to the original function if the order of the polynomial is increased, in other words, to have convergence between $P_n(x)$ and $f(x)$ In some cases this is true However this need not necessarily to be the case, like the Runges function (see Figure 5 9)

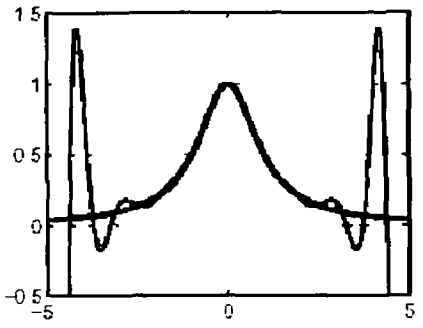

Figure 5 9 Interpolation by polynomials of degree 16 using equally spaced interpolation points for the Runges function $f = 1/(1 + 25x^2)$

5.4 Progressive Polynomial Calibration Implementation, Test and Validation

The progressive polynomial calibration method was simulated in SIMULINK A graphical diagram of the simulation is shown in Figure 5 10[27]

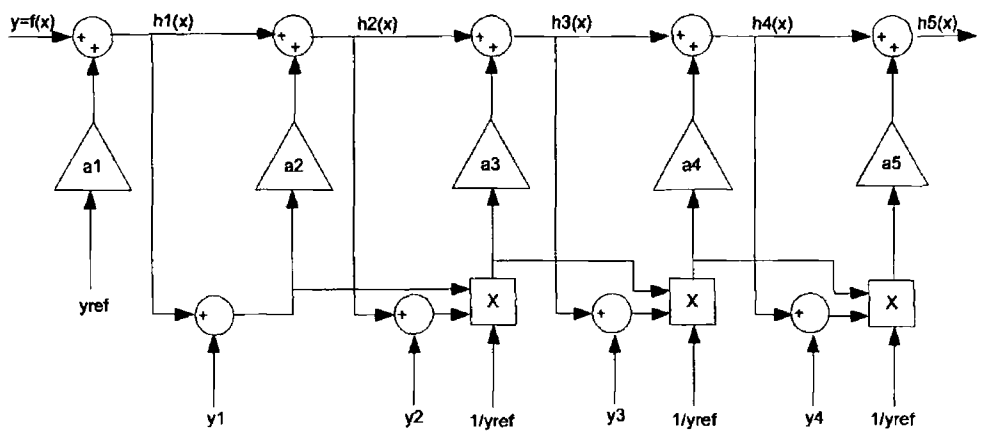

Figure 5 10 Graphical diagram of the polynomial calibration method

The simulation was quite straightforward In this simulation diagram the calibration coefficients are pre-calculated in C and then used by SIMULINK for testing Prior to this, the algorithm for calculating the linearised sensor transfer function $h_n(x)$ and calibration coefficients $a_n(x)$ was implemented in C The algorithm

consists of implementing equations 4.23 and 4.24 (repeated here).

$$a_n = \frac{y_n - h_{n-1}(x_n)}{y_{ref}} \prod_{i=1}^{n-1} \left(\frac{y_{ref}}{h_1(x_n) - y_i} \right)$$

$$h_n(x) = h_{n-1}(x) + a_n y_{ref} \prod_{i=1}^{n-1} \left(\frac{h_i(x) - y_i}{y_{ref}} \right)$$

Again, the compiled C code was tested for several different functions to validate accuracy. The same test functions were used, $\sqrt{x}$, x^{-1}, $\log x$ and e^x (Figure 5.11). All simulation results can be found in Appendix E.

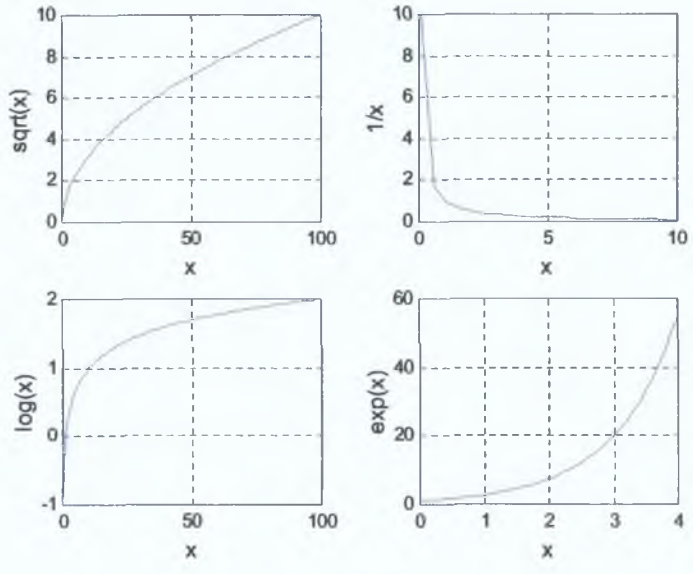

Figure 5.11: Test functions used for simulation.

The functions $\sqrt{x}$ and $\log x$ yielded reasonably good results. Difficulties arose when e^x and x^{-1} were tried. The simulation was performed, but the results gained were far from expected.

The calibration coefficients for the function e^x are calculated at $x = 0, 4, 2, 3$ and 1. Taking a closer look at Figure 5.12 it can be seen that the value of $h_5(x)$ at the point $x = 3$ does not match the desired straight line $g(x)$ even though the calibration coefficients are calculated correctly.

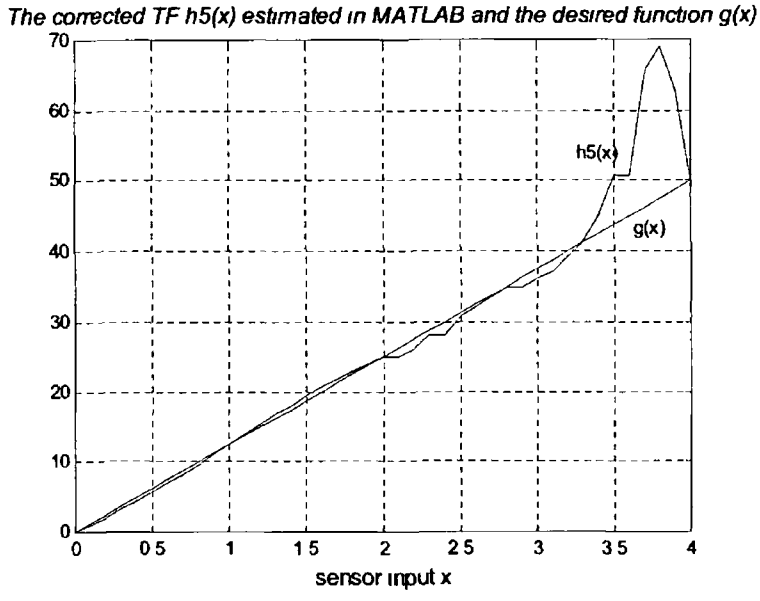

Figure 5 12 Estimated corrected sensor TF $h_5(x)$ in MATLAB for the sensor TF e^x using progressive polynomial calibration method in 5 points and desired sensor TF $g(x)$ for the calibration measurements taken in the following order $x = 0, 4, 2, 3$ and 1

The estimation of the corrected transfer function $h_5(x)$ in any point is made in MATLAB (running the simulation in SIMULINK) and the C code Figure 5 12 represents estimation in SIMULINK, whereas Figure 5 13 shows the results obtained in C code and then plotted in MATLAB

It is observed that in latter case estimation at the calibration points matches the desired transfer function and also at the point $x = 3$ The unexpected result in the former case is due to the MATLAB precision and the way functions and operations are defined in MATLAB software However the results obtained in C are more important as the C code is the part of the embedded system, a microcontroller, actual calibration and linearisation is carried out in C The errors corresponding to these two cases are shown in Figure 5 14 and Figure 5 15

The error in various non-periodic intervals on the full range is very large and it is much larger then the error produced using the Taylor polynomial interpolation Practically this amount of error is unacceptable It has been observed that the

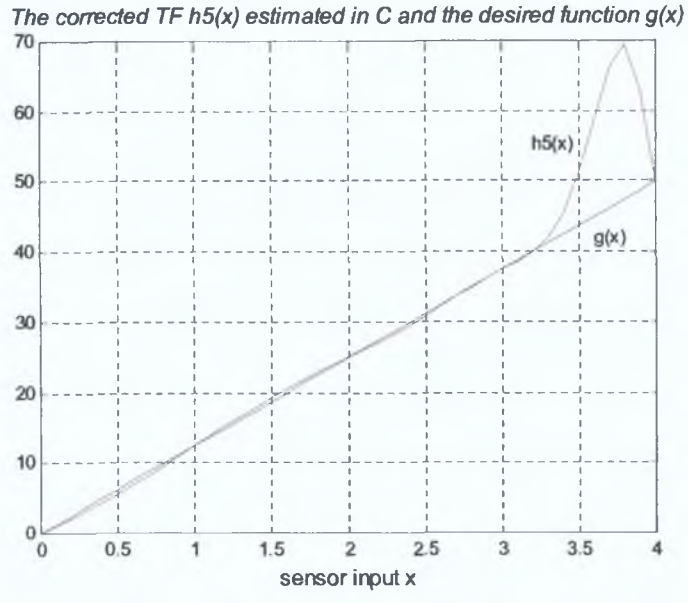

Figure 5.13: Estimated corrected sensor TF $h_5(x)$ in C for the sensor TF e^x using progressive polynomial calibration method in 5 points and desired sensor TF $g(x)$ for the calibration measurements taken in the following order $x = 0, 4, 2, 3$ and 1.

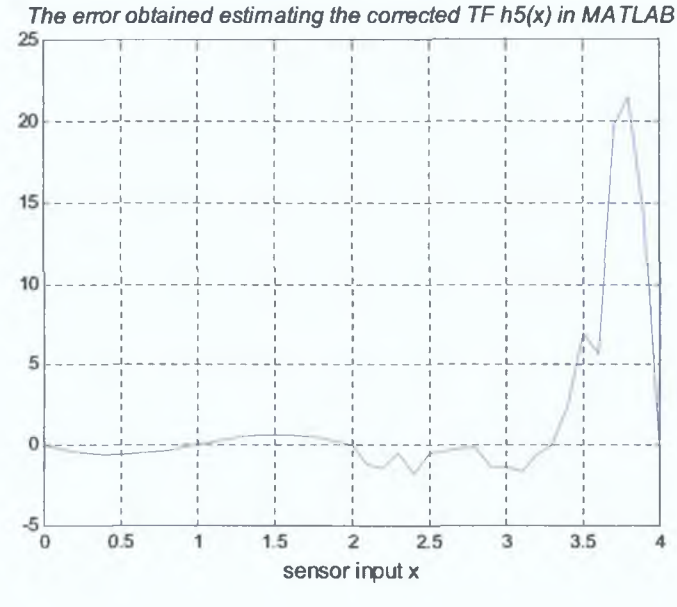

Figure 5.14: The error obtained estimating the corrected transfer function $h_5(x)$ in MATLAB for the function e^x in the points $x = 0, 4, 2, 3$ and 1.

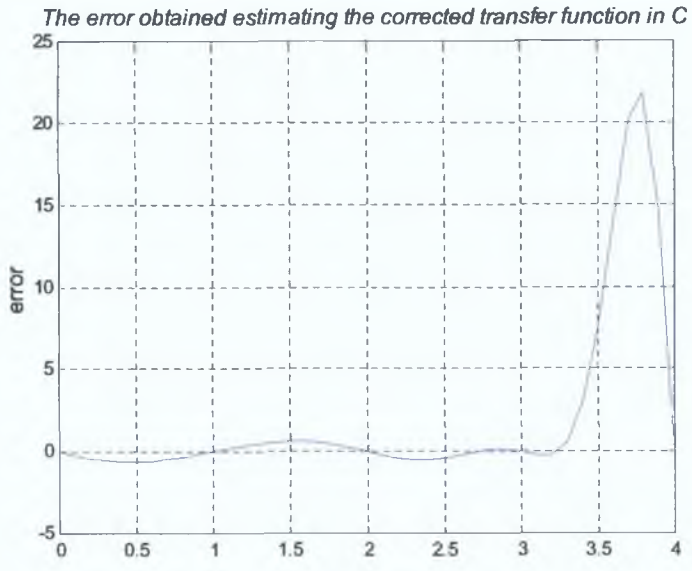

Figure 5.15: The error obtained estimating the corrected transfer function $h_5(x)$ in C for the function e^x in the points $x = 0, 4, 2, 3$ and 1.

error is unequally distributed over the full range for the other functions as well and that is practically negligible for the function $\sqrt{x}$ while for $\log x$ it is less than 3.5% (see Appendix E). These tests are run for 5 calibration measurements. Some of them are taken in equally spaced points based on rule 3 in choosing points. It is proved that this is not good enough if the function changes rapidly (like x^{-1} or e^x) and a much better way of choosing calibration measurements is to choose calibration points more closely in the region where a function "suffers" from these severe changes. Function x^{-1} shows a huge error that can be minimized by selecting calibration measurements carefully and introducing additional measurements. However, choosing calibration points in this case would be based on some knowledge about the sensor transfer function, and experience.

The next step was to test the functions e^{-x} and e^{2x} and see if the reason for such a large error for the function e^x was due to the rapid changes. The function e^{-x} changes more slowly on the chosen interval in contrast to e^{2x} that experiences more rapid changes than e^x. The example of corrected sensor transfer function for sensor transfer function described by e^{-x} is given by the Figure 5.16.

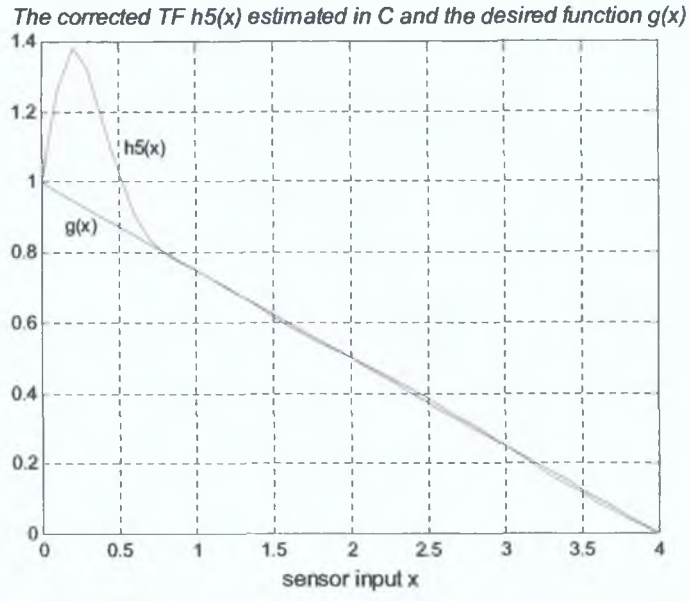

Figure 5.16: Estimated corrected sensor TF $h_5(x)$ in C for the sensor TF e^{-x} using progressive polynomial calibration method in 5 points and desired sensor TF $g(x)$ for the calibration measurements taken in the following order $x = 0, 4, 2, 1$ and 3.

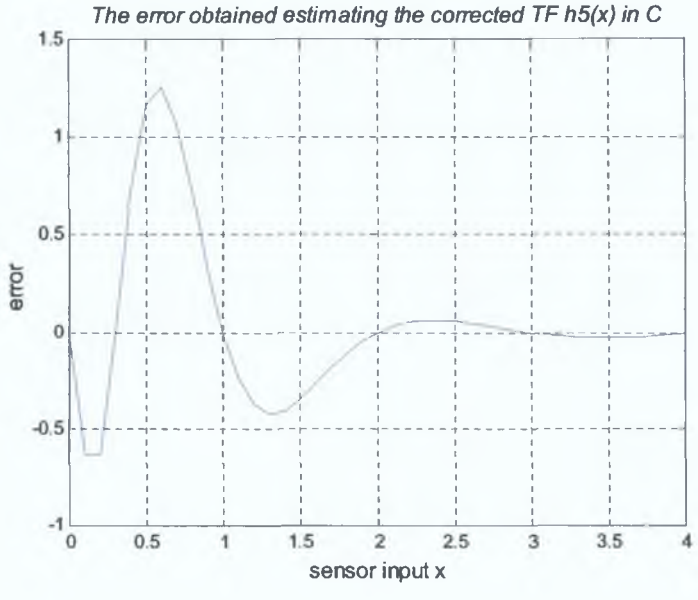

Figure 5.17: The error obtained estimating the corrected transfer function $h_5(x)$ in C for the sensor TF e^{-x} in the points $x = 0, 4, 2, 3$ and 1.

Examination of figures 5.17 and 5.18 confirmed the assumption that a large error of the exponential function is due to its rapid change on the chosen interval. Also conducting the error analysis (see Appendix E) on the other test functions (Figure 5.11) it was discovered that the error increases tremendously at the points where the slope (tangent) is greater than 0.8. However, the error of magnitude 10^6 as presented by the Figure 5.18 occurs on the interval were the function is not well behaved and usually the operating range of the sensor is restricted to the range where its response is well behaved.

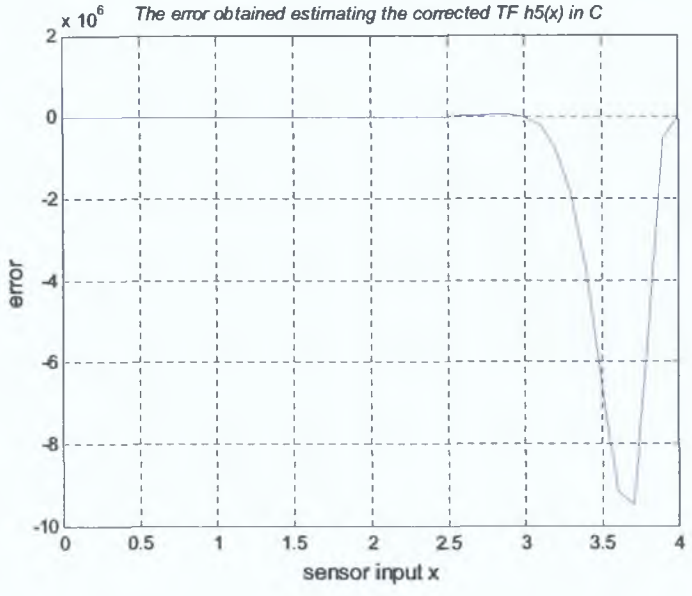

Figure 5.18: The error obtained estimating the corrected transfer function $h_5(x)$ in C for the sensor TF e^{2x} in the points $x = 0, 4, 2, 3$ and 1.

The order of taking samples, i.e. taking the fourth sample x_4 to be either $x = 3$ or $x = 1$, on half of the intervals did not indicate any difference in the results for the functions $\sqrt{x}$ and $\log x$, while the results for the function e^x differ greatly. The results again differed at the point $x = 3$ for the evaluation in MATLAB and C for the same reason. Only results obtained in C are shown here (Figure 5.19) as they represent the real case (what would actually happen if the algorithm is implemented on a microcontroller).

The error in this case is even bigger than when the measurements are taken in

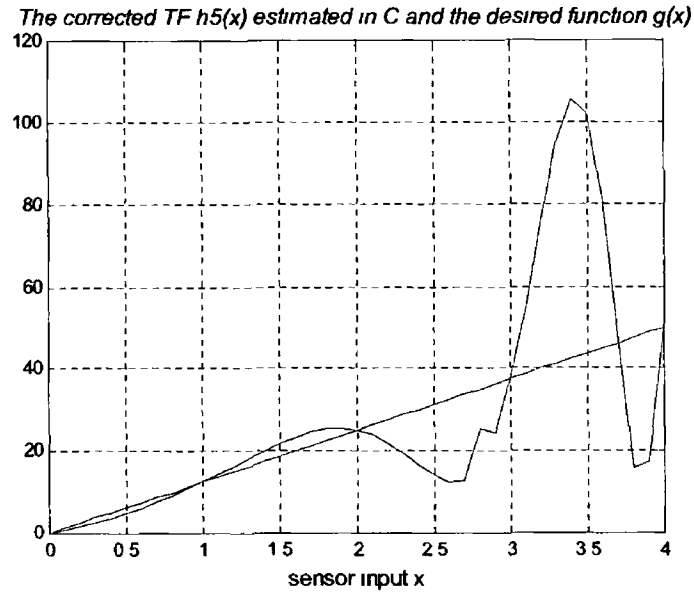

Figure 5 19 Estimated corrected sensor TF $h_5(x)$ in C for the sensor TF e^x using progressive polynomial calibration method in 5 points and desired sensor TF $g(x)$ for the calibration measurements taken in the following order $x = 0, 4, 2, 1$ and 3

the order $x = 0, 4, 2, 3$ and 1 From the results obtained a connection between an error and the order of taking calibration measurements can be found Where the error was very small ($\sqrt{x}$ and $\log x$) the results illustrated no difference when the order of taking measurements was changed The large error always occurred when the results were affected by changing the order of taking measurements

Contrary to the previous example when the sensor transfer function was described with e^x, the error is less when the calibration measurements are taken in the order $x = 0, 4, 2, 1$ and 3 (Figure 5 20) The conclusion that can be drawn is that the effect on the amount of the error of taking calibration measurements in certain order depends on the function itself As it depends on a function itself, it is impossible to predict it and hence correct in some way in software

The same diode circuit that had been used for Taylor series validation is used for validation of progressive polynomial calibration (Figure 5 6) The corrected transfer function is shown in Figure 5 21

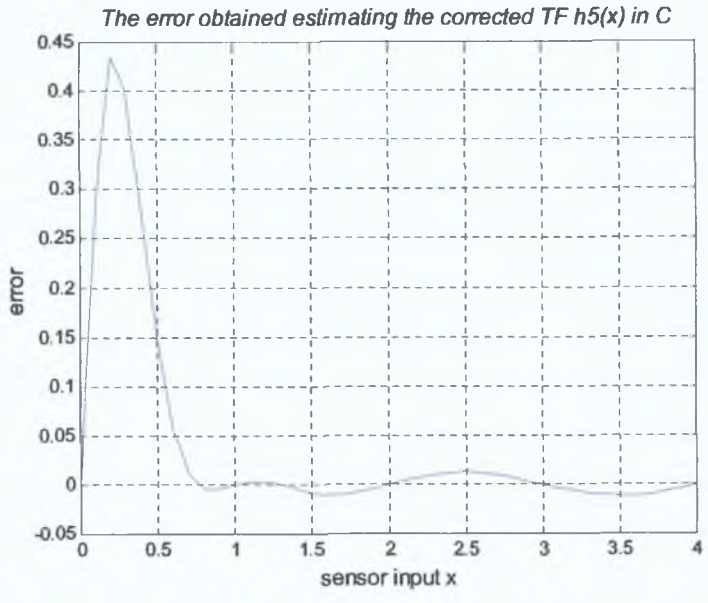

Figure 5.20: The error obtained estimating the corrected transfer function $h_5(x)$ in C for the sensor TF e^{-x} in the points $x = 0, 4, 2, 1$ and 3.

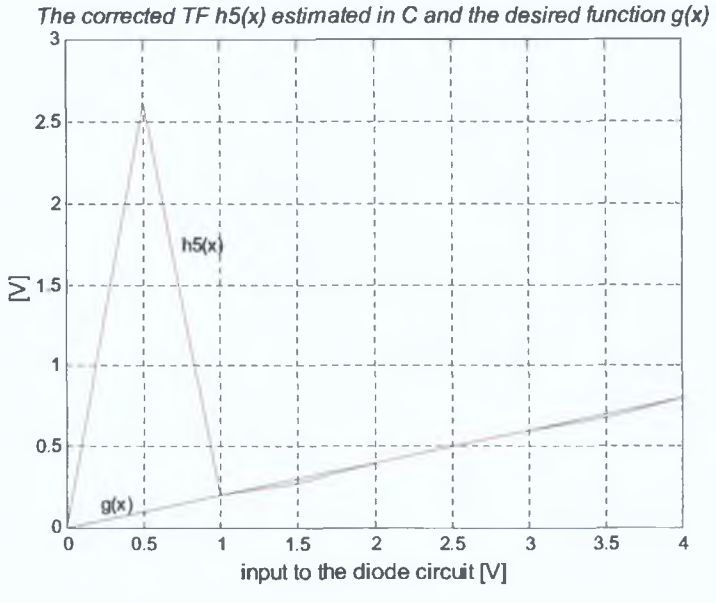

Figure 5.21: Corrected sensor TF $h_5(x)$ of the diode circuit estimated in C, using progressive polynomial calibration method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB.

A large error on the interval $x = [0, 1]$ was expected as the circuit experiences fast changes in the output on the specific interval In line with the previous conclusions, tests showed large differences in the results when the order of taking measurements was changed

5.4 1 Summary

For the functions that change rapidly (when applying progressive polynomial calibration method), the order of using measurements is very important as well as choosing calibration measurements Comparing results for e^x and e^{-x} it can be seen that when observing the order of using calibration measurements it is better to take the fourth calibration measurement where the function changes rapidly In terms of choosing calibration measurements for the functions that change rapidly it is better to "squeeze" calibration measurements in the affected (by the large error) interval of the full range This should be done by placing the third and rest of the measurements in the affected area, keeping the first and the second measurement at both ends of the full range for offset and gain correction It is also discovered that when taking the rest of the measurements one of them should be at half of the range However, this measurement on the half of the range gives better results when it is taken as a fourth measurement, taking as the third measurement point between half of the range and the last measurement on the affected interval Example of this is shown for x^{-1} in Appendix E (Figure E 3) Further improvement can be achieved taking additional measurements and/or further analysis of the sensor transfer function in order to adjust the calibration measurements accordingly The larger a range the more points for severe functions are needed in order to minimize the errors

5.5 Tests on Actual Sensor Responses

Earlier in this chapter, investigated functions represent extreme cases in which a huge error is expected In this section Taylor series, Gaussian elimination,

Lagrange interpolation and progressive polynomial calibration are implemented for two functions that represent actual sensor responses Function f_1 describes the Humirel HS1100/HS1101 capacitive RH sensor [48] whereas function f_2 describes the Figaro TG832 semiconductor gas sensor [49] These responses are shown in Figure 5 22

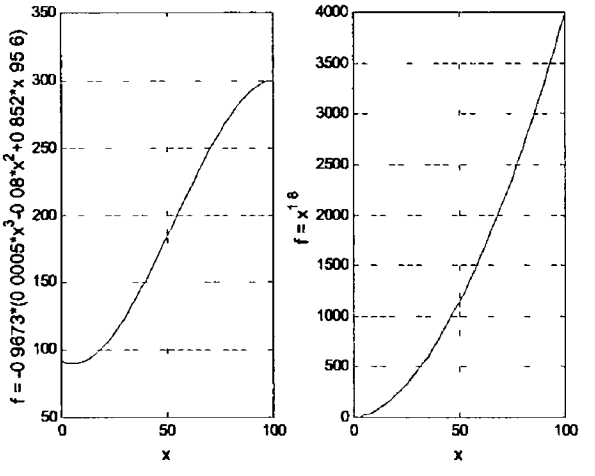

Figure 5 22 Sensor responses used as test functions described with

$$f_1 = -0\ 9673(0\ 0005x^3 - 0\ 08x^2 + 0\ 852x - 95\ 6) \text{ and } f_2 = x^{1\ 8}$$

5.5 1 Taylor Series

Based on the test results from the previous section the piecewise Taylor series polynomial for both of the functions f_1 and f_2 (as shown in Figure 5 22) were evaluated by the rule represented by the equation 5 7 The results obtained demonstrate very good approximation of the original functions with the error less than 1 5% of full scale (FS), where FS is the function response to full (Figure 5 23)

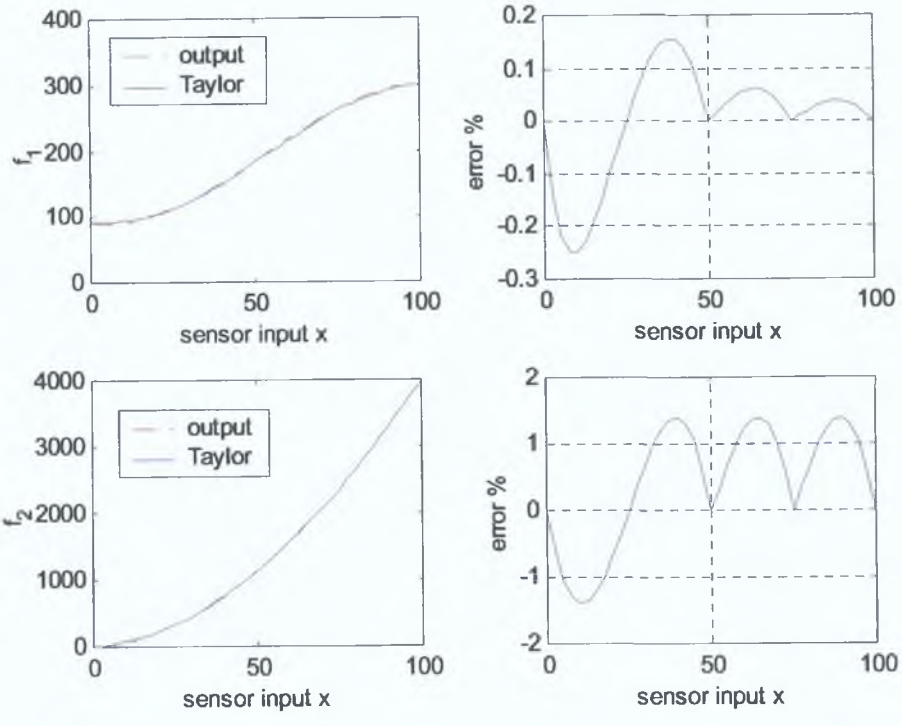

Figure 5.23: Taylor polynomial interpolation and error percentage for the functions f_1 and f_2.

5.5.2 Gaussian elimination

To construct the third order polynomial using Gaussian elimination, 4 points are needed to solve the system of 4 equations (4 unknown calibration coefficients). Points are taken in such a way to get the better fit for the inverse function. The error for the functions f_1 and f_2 is about 11% of the FS (Figure 5.24), where the full scale is the difference between maximum and minimum value of the y-axis and it changes through out the thesis with the respect of the function evaluated.

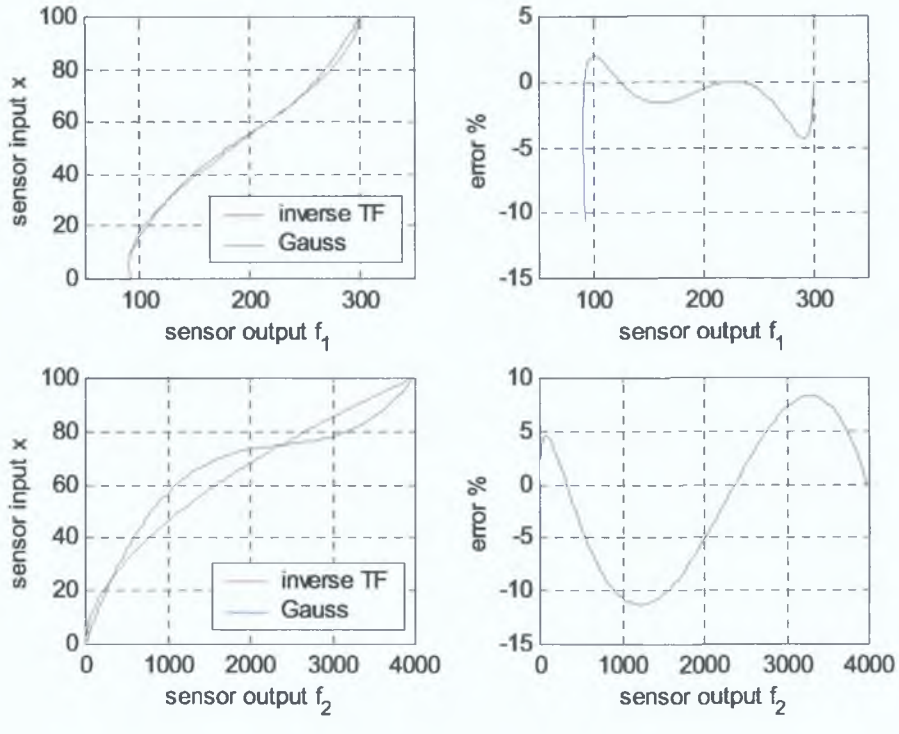

Figure 5.24: Interpolation with the third order polynomial using Gaussian elimination for calculating the coefficients and error percentage for the inverse transfer functions of the functions f_1 and f_2.

5.5.3 Lagrange Interpolation

Lagrange interpolation for the same functions was restricted to 5 points to avoid polynomial oscillation common for higher degree Lagrange polynomials. Points are taken equally spaced on y axis (sensor output) as this was proven to give less error and less *oscillation*. The error obtained (Figure 5.25) is less than 6% of FS.

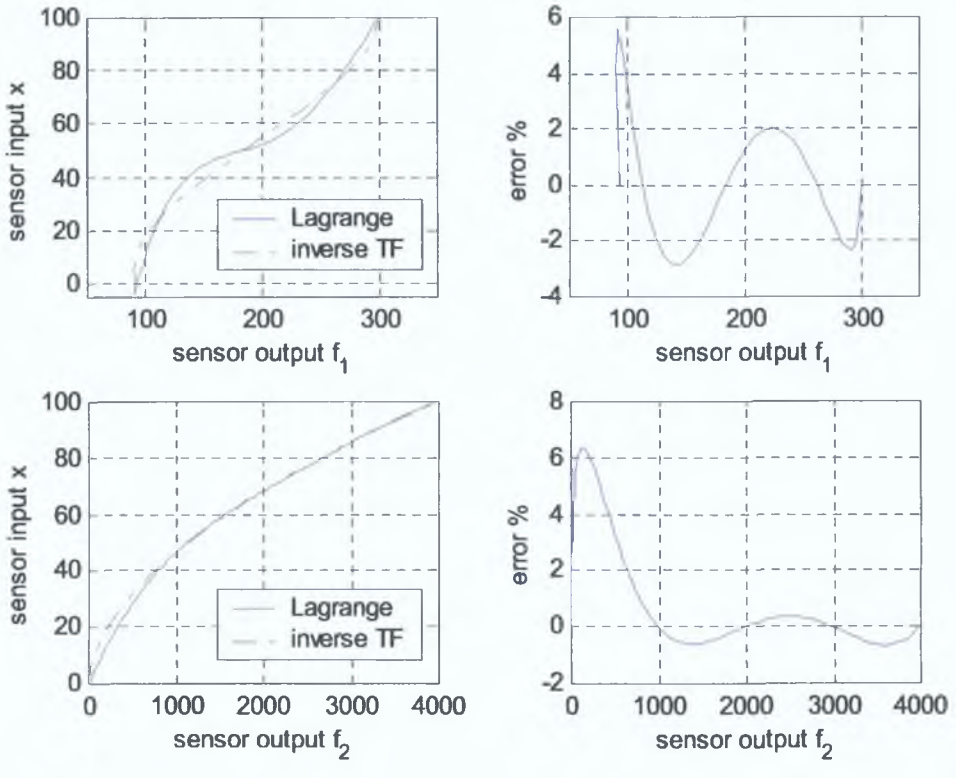

Figure 5.25: Lagrange polynomial interpolation and error percentage for the inverse transfer functions of the functions f_1 and f_2.

5.5.4 Progressive Polynomial Calibration

Simulation and implementation in C for progressive polynomial calibration showed that the error is large for both functions as they both suffer from rapid changes. The results for f_1, corrected transfer function and corresponding error for 5 calibration measurements, are shown in figures 5.26 and 5.27.

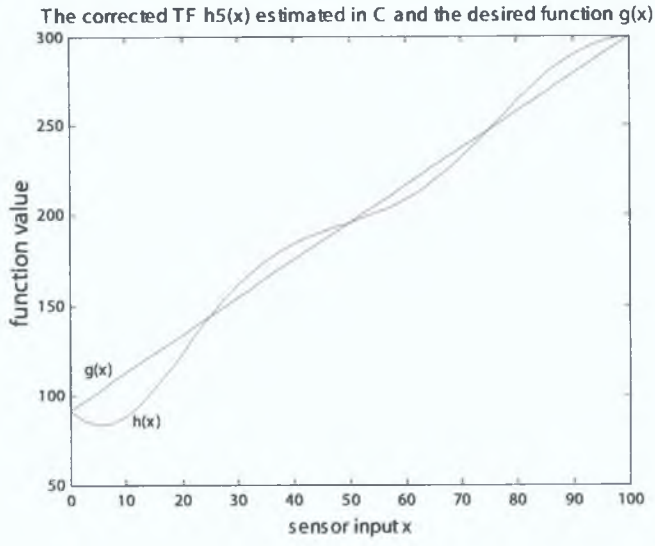

Figure 5.26: Corrected sensor TF $h_5(x)$ of the sensor TF f_1 using the PPC method with 5 points and the desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 100, 50, 75, 25\}$).

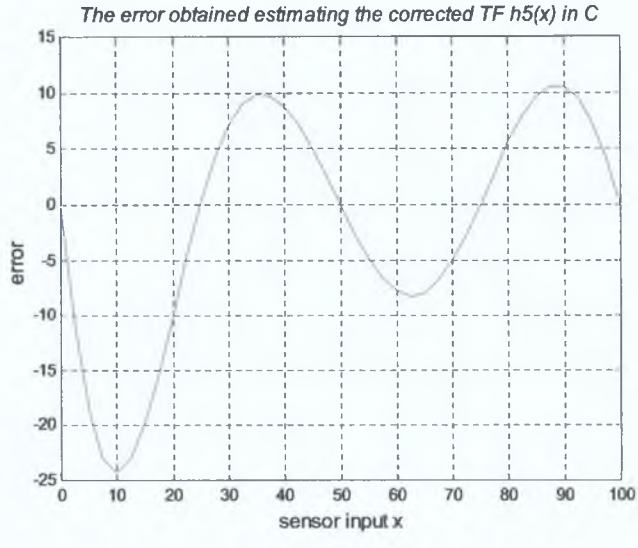

Figure 5.27: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0, 100, 50, 75, 25\}$).

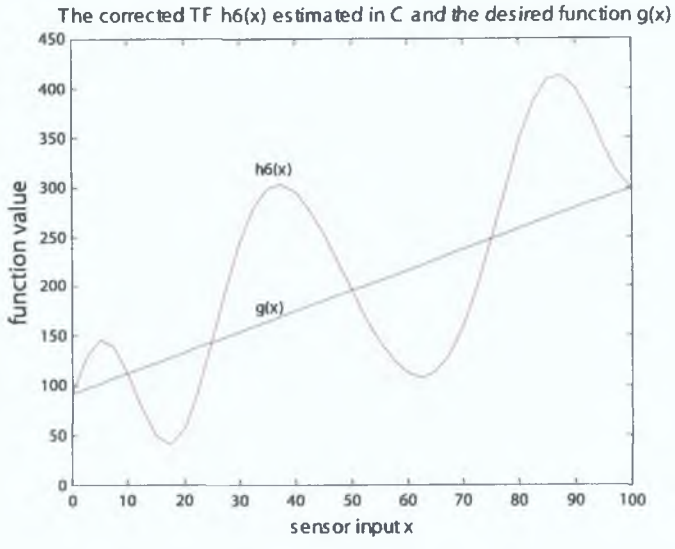

Figure 5.28: Corrected sensor TF $h_6(x)$ of the sensor TF f_1 using the PPC method with 6 points and the desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 100, 50, 75, 25, 10\}$).

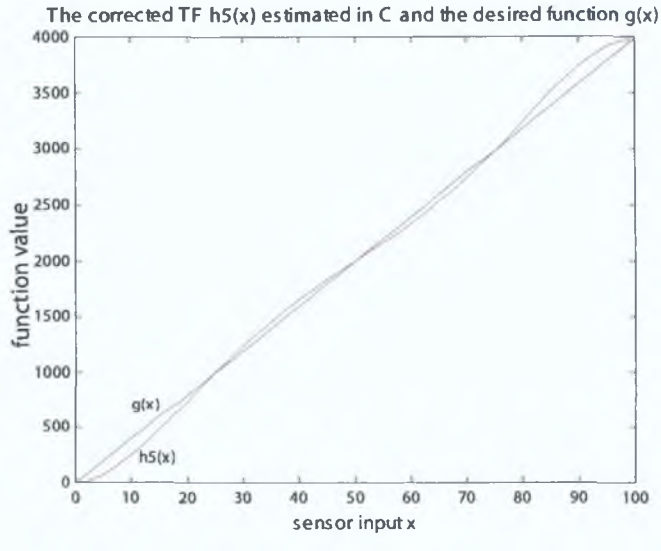

Figure 5.29: Corrected sensor TF $h_5(x)$ of the sensor TF f_2 using the PPC method with 5 points and the desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 100, 50, 75, 25\}$ - equally spaced points on x axis).

The measurements are chosen equally spaced on x axis (sensor input) It can be seen that the error is as large as 12% of FS In order to reduce the error an additional measurement is taken at the point $x = 10$ where the *oscillation* was the most pronounced It was expected that this would reduce the error as in tests before but, on the contrary it produced even greater *oscillations* of the polynomial and larger error (Figure 5 28)

Two tests were run for the function f_2 One test was when calibration measurements were taken as equally spaced points on the x axis (sensor input) and the other test was for equally spaced points on the y axis (sensor output) The results are presented in Figure 5 29 (equally on x axis) and Figure 5 30 (equally on y axis)

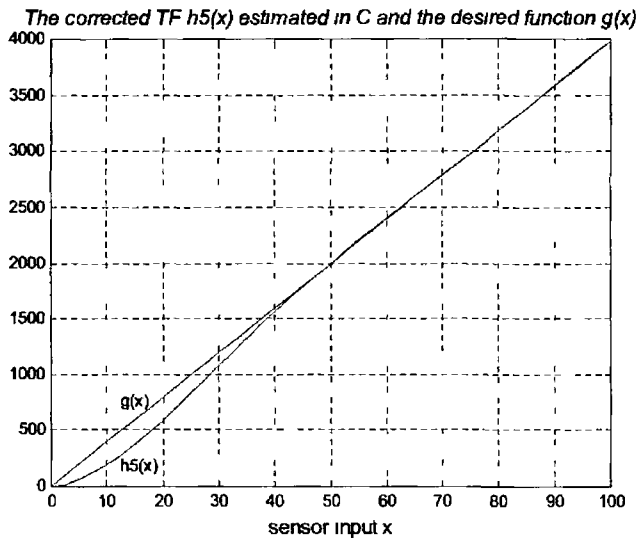

Figure 5 30 Corrected sensor TF $h_5(x)$ of the sensor TF f_2 using the PPC method with 5 points and the desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 100, 68\ 04, 85\ 23, 46\ 294\}$ - equally spaced points on y axis)

Contrary to what was expected, changing the order of taking the two last measurements in the latter case did not have any effect whatsoever The error in this case is less than 6% of FS, while in former case is less than 4% of FS

5.6 Summary and Conclusions

Different linearisation methods have been examined Taylor series, Gaussian elimination, Lagrange interpolation and PPC methods gave the least error and the best results The Taylor series requires equally spaced points which in practice is almost impossible to achieve (for example to set the gas concentration at the exact level) Gaussian elimination, Lagrange interpolation and PPC are more flexible On the other hand these calibration methods require some knowledge of the sensor transfer function and the selection of measurement points

All four methods were tested on the third and second order polynomials that represent responses for vast variety of sensors (see Figure 5 22)

Taylor series showed very good results for 5 calibration measurements evaluated in 3 middle points by the rule that has been devised and is described by equation 5 7 The error in these cases was less than 0 3% of full scale for f_1 and 1 5% of full scale for f_2

Gaussian elimination produced an error of 11%

Lagrange interpolation when applied, gave fairly good results, the error less than 6% of full scale, but also produced and showed inevitable polynomial *oscillation* associated with Lagrange interpolation

Progressive polynomial calibration produced an error less than 6% of full scale This amount of error is associated with the polynomial oscillation

When extreme cases are inspected (functions $\sqrt{x}$, x^{-1}, $\log x$ and e^x) all methods *experienced* difficulties and the errors were much larger These large errors were usually on the part of the range where functions change rapidly while on the rest of the range errors were small However, the operating range of a sensor is normally restricted to the range where its response is well behaved

In addition to these tests, some other functions and methods are tested and summarized in the Table 5 6 The methods that are used for testing and comparison and are not explained in this thesis are piecewise linear, piecewise quadratic and least squares quadratic and cubic approximation [38]

Taylor series was abandoned because of the nature of the method and its need for equally spaced points which in practice is almost impossible to achieve. Even a small digression causes a large error. Lagrange interpolation and progressive polynomial calibration share a common problem: how to choose calibration measurements to get the best approximation and avoid the global dependence on local properties (polynomial oscillation). However, progressive polynomial calibration showed good results for certain functions and it was tested further.

FUNCTION	METHOD (ERROR [%])								
	GAUSS 2nd order of inv	*PW-linear of inv*	*PPC*	*GAUSS 3rd order of inv*	*LS-quadratic of inv*	*PPC*	*PW-quadratic* of inv*	*LS-cubic of inv*	*PPC*
$f1 = -0.9673(0.0005x^3 - 0.08^2 + 0.852x - 95.6)$	12%	15.00%	14.40%	12.00%	12%	12.00%	12%	12%	12%
$f2 = x^{2.5}$	20.60%	15.80%	30.00%	15.84%	14.30%	14.00%	15.85%	11.10%	8.00%
$f3 = x^{1.8}$	11%	14.80%	10%	7.75%	8.50%	5.50%	7.75%	5.42%	4%
$f4 = x^{1/2.5}$	2.50%	6%	4.80%	0.20%	2%	1.67%	0.40%	0.18%	0.67%
$f5 = x^{1/1.8}$	1%	7.70%	1.23%	0.20%	0.62%	0.33%	0.25%	0.20%	0.25%
NUMBER OF POINTS	**3**	**3**	**3**	**4**	**4**	**4**	**5**	**5**	**5**
NUMBER OF COEFFICIENTS	**3**	**4**	**3**	**4**	**3**	**4**	**6**	**4**	**5**

Table 5.4: Summary of different methods.

* Gaussian elimination is used for calculating the coefficients in PW-quadratic method.

Note: All the functions are tested on $x = 0 - 100$ scale.

Colours indicate which method yielded the smallest error.

Different colours symbolize a different number of the calibration points.

Chapter 6

Compensation Techniques

Many sensors have outputs that are nonlinear and that are sensitive to temperature and humidity In these cases linearisation, temperature compensation and humidity compensation are required Linearisation, temperature and humidity compensation can be performed by a processor

Analogue calibration of sensors and compensation of temperature effects or nonlinearities often make sense before the sensor signal is applied to an ADC The disadvantage of analogue compensation techniques is that one can rarely perform the compensation in more than one dimension (i e temperature and humidity) The digital approach, which combines the use of numerical methods and standard data acquisition techniques, results in efficient calibration and evaluation techniques

Digital compensation can be done either using a look-up table or some kind of compensating algorithm resulting in the real time compensation The speed of such algorithm is very good, since it does not require recalculation

The main disadvantage of using look-up tables is large memory consumption which is often unaffordable in embedded sensor systems Small amounts of memory limit the use of digital sensor compensation by use of look-up table approach

Real time compensation is automatic compensation controlled by software to continuously adapt to various environments yielding consistent sensor performance

6.1 Design of a Compensation Module

Many sensors are susceptible to temperature and humidity changes In these cases the sensors' readings will be accurate only for the temperature or humidity at which it was calibrated Automatic compensation is much easier to implement than to control temperature and humidity during sensor operation Therefore compensation is essential The temperature and humidity sensors used for compensation have to be mounted in close proximity to the sensor for which readings need to be compensated

The compensation is performed before the linearising calibration and compensation coefficients are stored in the microprocessor memory This way there is no need to recalibrate the sensor as the ambient conditions varies The compensation and linearising calibration process is represented by Figure 6 1, where x is the sensor output, v_n is compensated sensor output by the nth variable and v_{lin} is corrected sensor output

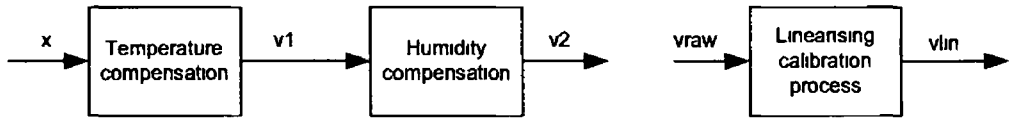

Figure 6 1 Block diagram of the compensation and linearising calibration process

Some sensors have an output characteristic that has a fairly linear dependence on temperature and humidity while for others it is non-linear For those that have a linear dependence, calculation of compensation coefficients is nothing but solving the coefficients of the line equation The line equation requires the calculation for only two compensation coefficients and therefore two measurements are sufficient for compensation of one variable (such as temperature, humidity or pressure)

If a sensor characteristic is given as in the Figure 6 2, then it can be seen that for the same sensor input x_{ref}, the output varies with temperature variations

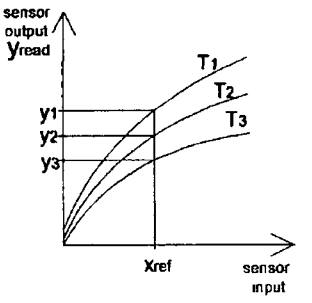

Figure 6 2 Variations of the sensor output by temperature changes

Assume that this variation can be expressed by the line equation 6 1 (Figure 6 3)

$$y_T = a_1 + a_2 T \quad \text{for } x_{ref} = const \tag{6 1}$$

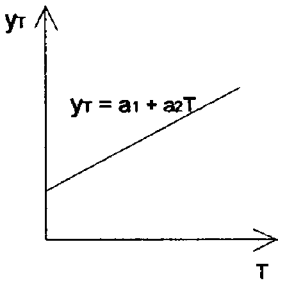

Figure 6 3 Temperature compensation

Solving the equation 6 1 yields

$$a_2 = \frac{y_{read} - y_{ref}}{T_{read} - T_{ref}} \tag{6 2}$$

and

$$a_1 = y_{ref} - a_2 T_{ref} \tag{6 3}$$

where y_{ref} is the reading taken at reference temperature T_{ref} and y_{read} is the measurement taken at some temperature of the measurement T_{read} Combining the equations 6 1 and 6 3 it is possible to express the compensated value as

$$y_{comp} = y_{read} - a_2(T_{read} - T_{ref}) \tag{6 4}$$

which means it is enough to store only one compensation coefficient a_2 and save the memory space y_{comp} represents the compensated sensor output and approaches to y_{ref} after compensation and the difference between these two values corresponds to the error of the compensation

The same method can be used for humidity compensation as well As it can be seen in the Figure 6 2 sensor output dependence on temperature can vary for different input values To improve the accuracy of compensation coefficients several measurements can be taken for different sensor inputs at different temperatures and then the mean value for the coefficient is taken The other way would be to calculate the compensation coefficient for the T_{ref} at the mid-scale as it is presented in the Figure 6 2

Normalizing the compensated sensor output (Figure 6 4) for each variable it is possible to get the correction factor and then in order to get the corrected sensor output y_{corr} it is necessary to multiply the output read by the sensor y_{read} with the correction factor y_T/y_{ref}

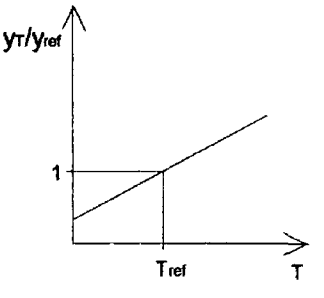

Figure 6 4 Correction factor

If the sensor output changes in non-linear manner with changes of a compensation parameter (temperature, humidity) then the equation of a line is not good enough as the function between the output and the compensated variable is a higher order polynomial To calculate the calibration coefficients in this case, one of the already investigated methods, Gaussian elimination for example, can be used thus saving the microcontroller storage memory if the Gaussian elimination method is also used in the linearising calibration process

Chapter 7

Test and Validation

7.1 Design of the Prototype Electronics for the Calibration and Compensation Module

For the test of a calibration and compensation module, a Figaro TG832 semiconductor gas sensor and a Sensirion SHT11 temperature and humidity IC sensor are used (Figure 7.1).

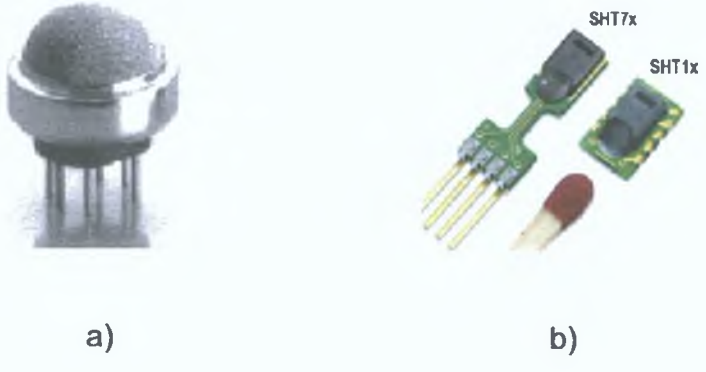

Figure 7.1: a)Figaro TG832 semiconductor gas sensor; b) Sensirion SHT11 temperature and humidity sensor.

The Figaro TG832 gas sensor is chosen due its nonlinear characteristics and temperature and humidity dependence. It is sensitive to R134A refrigerating gas, used in this experiment. SHT11 includes two calibrated temperature and humidity sensors. As the accuracy of the relative humidity sensor depends on temperature, humidity readings are also self compensated by temperature. Humidity and temperature are simultaneously read into the system.

The Microchip PIC microcontroller performs continuous real time temperature and humidity compensation as well as the linearisation based on calculated compensation and calibration coefficients.

For test purposes the TG832 gas sensor and SHT11 temperature and humidity sensor are built on separate boards and interfaced to the microcontroller electronics as presented in Figure 7.2.

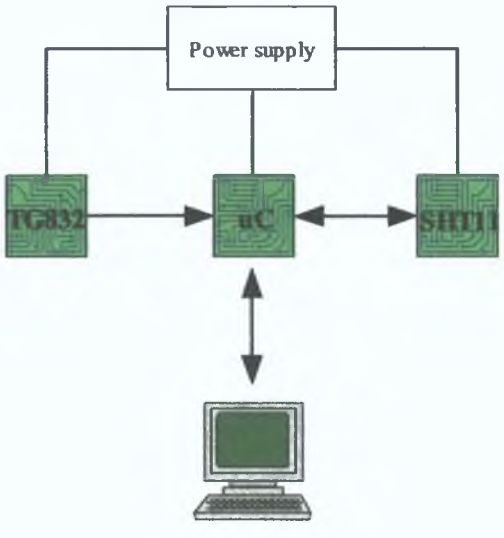

Figure 7.2: Conceptual diagram of the printed circuit boards used and their interfaces.

Also the prototype of the calibration and compensation module is assembled on a single board on which the sensors and the microcontroller are mounted together.

7.2 Code Description

The linearisation technique to be performed depends on the sensor characteristic Two methods are chosen and implemented on a PIC microcontroller for testing Gauss elimination and progressive polynomial calibration (PPC) methods The Gauss elimination method can also be used for calculating compensation coefficients where the relationship between the sensor output and the compensated parameter is not linear

The program loaded onto the PIC for testing purposes performs the calculation of the compensation coefficients first and then calculates calibration coefficients, which are both later on used for correcting the sensor output (see Figure 6 1)

The program is very simple For execution it is necessary to have very well controlled environmental conditions (temperature and humidity) in order to perform the calibration The idea is to change one variable while keeping the others constant In the case of the gas sensor, the gas concentration is kept constant and the gas reading is taken for different humidity readings while the temperature is constant These readings are then used to calculate the humidity compensation coefficients Then the process is repeated for the temperature compensation coefficients, keeping the humidity constant The flowchart of the sample program for compensation and calibration is shown in Figure 7 3

If linearisation is realised by interpolation, the Gauss elimination method is used to calculate the calibration coefficients of the inverse sensor transfer function and a computer algorithm is simple The computer algorithm for PPC implementation is slightly more complicated and it is presented by flowchart in Figure 7 4

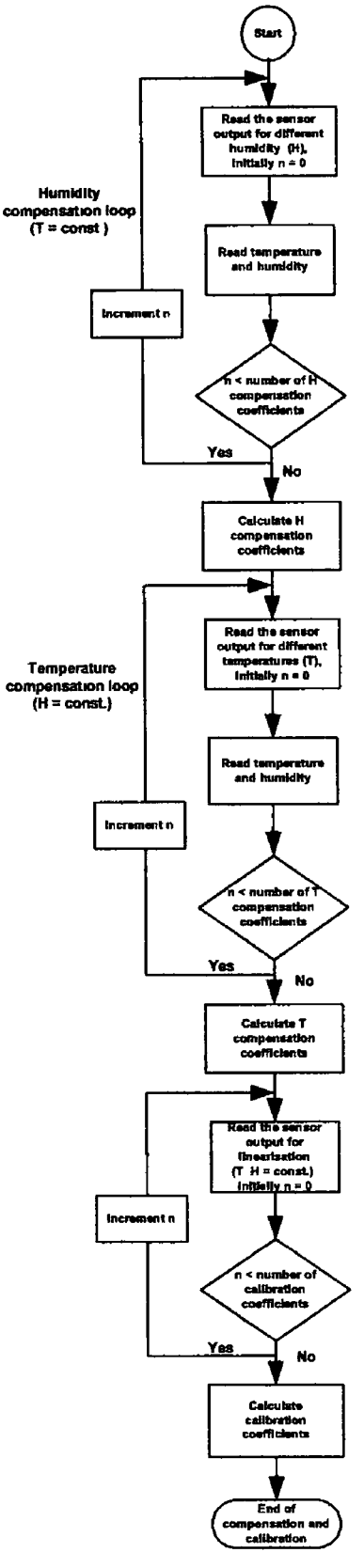

Figure 7 3 The flowchart of the sample program for compensation and calibration

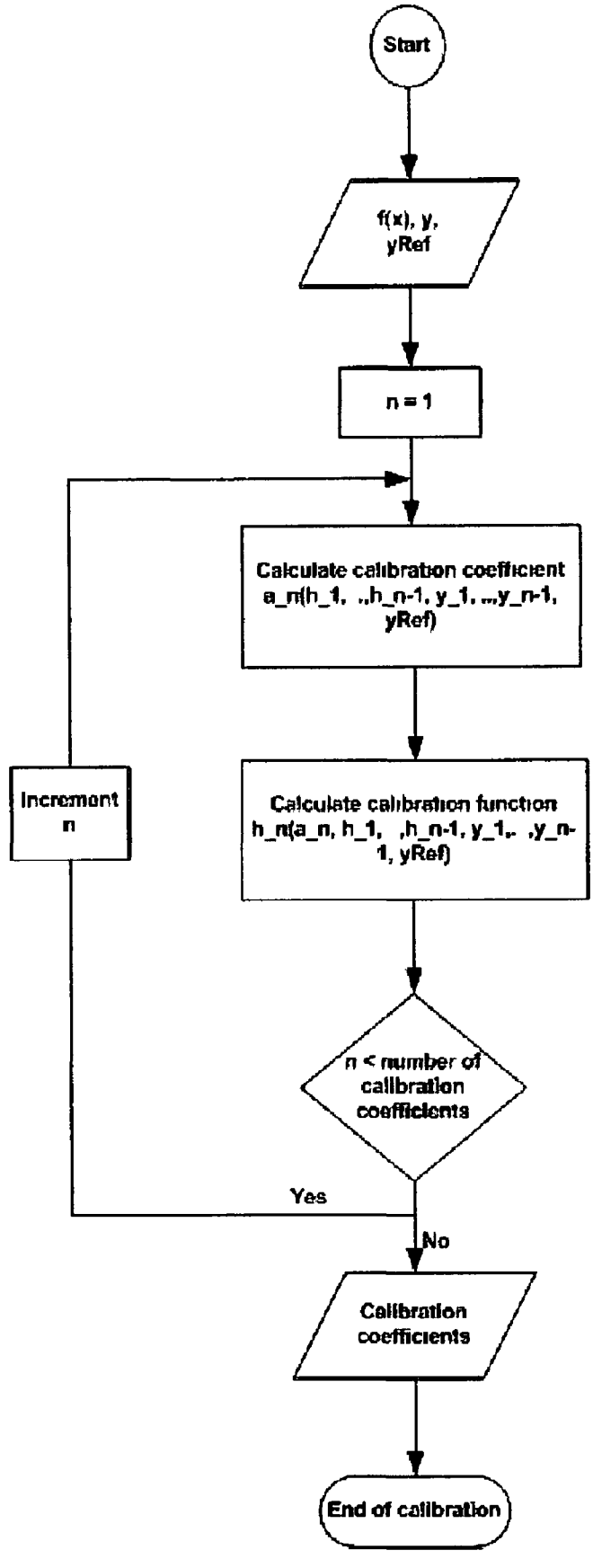

Figure 7 4 The flowchart of the PPC calibration

7.3 Experimental Procedure and the Test Rig Setting

Before starting any tests on the compensation and calibration module it was necessary to determine how the sensor output behaved with temperature and humidity changes since the data sheet provided from the manufacturer of the gas sensor did not provide sufficient information.

For that purpose it is decided to use a small box (Figure 7.5), pretty well sealed (of high ingress protection IP value) to prevent gas leakages (R134A refrigerating gas) and ensure the lid of the box would withstand the pressure from the gas bottle; the gas and the temperature and humidity sensors were mounted in close proximity within the box. Boxes with different IP values are tested (IP65, IP66, IP67 and IP68). The small dimensions of the box provide low gas intake and reduce the cost of the experiment. The test box with the sensors is then placed in the temperature controlled chamber to facilitate the different test temperatures. The hole on the box is made to let the air out while filling in the box with the gas.

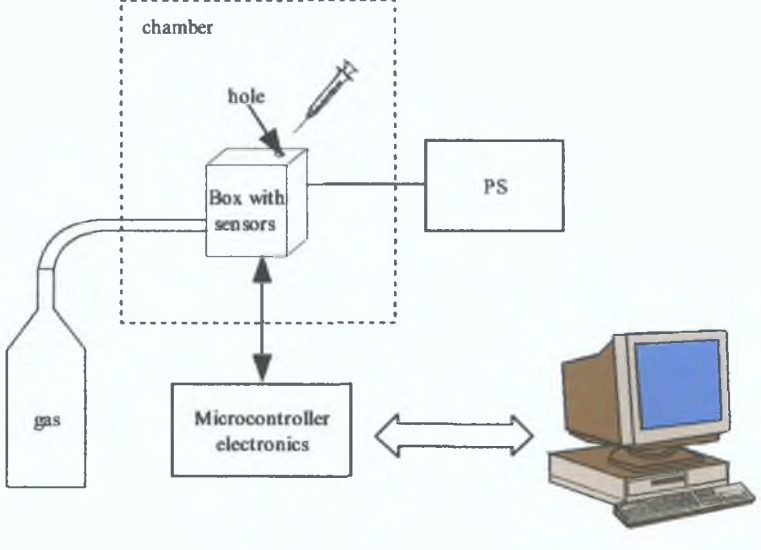

Figure 7.5: Experiment setting.

The humidity is varied keeping the gas concentration and the temperature con-

stant to obtain the gas sensor output changes with changes in the humidity and similarly for the sensor output changes with temperature, keeping the humidity and the gas concentration constant and varying the temperature

At first, attempts were made to control the humidity using different saturated salts solutions and keep the temperature constant A list of saturated salts and their constant humidity values is obtained from Humirel application note HPC020 "Basics of Relative Humidity Calibration for Humirel HS1100/HS1101 Sensors" and shown in Figure 7 6

Temperature°C	Lithium Chloride Solution $LiCl\ H_2O$	Magnesium Chloride Solution $MgCl\ 6\ H_2O$	Magnesium Nitrate Solution $Mg(NO_3)_2\ 6\ H_2O$	Sodium Chloride Solution $NaCl\ 6\ H_2O$	Potassium Chloride Solution $K\ SO_4$
5	1	33 6 ± 0 3	58	75 7 ± 0 3	98 5 ± 0 9
10	13	33 5 ± 0 2	57	75 7 ± 0 2	98 2 ± 0 8
15	12	33 3 ± 0 2	56	75 6 ± 0 2	97 9 ± 0 6
20	12	33 1 ± 0 2	55	75 5 ± 0 1	97 6 ± 0 5
25	**11 3 ± 0.3**	**32 8 ± 0 3**	**53**	**75 3 ± 0 1**	**97 3 ± 0 5**
30	11 3 ± 0 2	32 4 ± 0 1	52	75 1 ± 0 1	97 0 ± 0 4
35	11 3 ± 0 2	32 1 ± 0 1	50	74 9 ± 0 1	96 7 ± 0 4
40	11 2 ± 0 2	31 6 ± 0 1	49	74 7 ± 0 1	96 4 ± 0 4
45	11 2 ± 0 2	31 1 ± 0 1	/	74 5 ± 0 2	96 1 ± 0 4
50	11 1 ± 0 2	30 5 ± 0 1	46	74 6 ± 0 9	95 8 ± 0 5
55	11 0 ± 0 2	29 9 ± 0 2	/	74 5 ± 0 9	/

Figure 7 6 Relative humidity of saturated salt soultions

Varying the humidity with the salts did not give expected results, i e the humidity did not agree with the value presented by the table (Figure 7 6) and it took a long time to settle and stay constant during the experiment, due possibly to porosity of the material the box is made of (humidity tends to reach the humidity outside the box over time) Also the salts were damaging the PCB (printed circuit board) tracks and components as the salt tended to "creep" up on the walls of the box

Having this in mind alternative methods of varying the humidity were examined and the method that gave the best results was injecting drops of water through the small hole with a syringe (Figure 7 5) In this way it was possible to keep the temperature constant and vary the humidity However, varying the humidity this

way it was not possible to control it as the settling time depends of the amount of the water injected and the starting %RH value was different each time depending on the %RH in the room Also a "jump" in the humidity reading was experienced when water was injected and that jump also depended on the amount of water injected into the box To get around these problems, silica gel was put into the box to lower the humidity inside the box and to get approximately the same starting %RH reading by quickly removing it from the box before injecting in the water Nevertheless, the results were not satisfactory

While trying to characterise the gas sensor output other difficulties were also encountered These difficulties were associated with preventing the leakages, that is to say finding a way to seal the holes on the box made for the power supply and communication interface Different types of glues were used in these purposes and the results showed that the TG832 semiconductor gas sensor detects glues (even dry), although it was targeted for refrigerating gases This resulted in a high reading even if no gas was present

At this stage different boxes were also examined trying to determine which one was the least porous Boxes of different materials were tested including polystyrene (PS), polyester (GRP), polycarbonate (PC), aluminium diecast and glass The Al diecast box demonstrated better behaviour than the others The other advantage of the aluminium box was that some of the plastic boxes were painted causing the detection of the paint by the gas sensor

Moreover an increase of the gas reading was detected over the time for the constant temperature and humidity readings as if the gas sensor was "burning" something inside the box (such as oxygen) which wasn't investigated further Presuming that this may have had to do something with the small constrained space of the box, the same test was repeated with a box of bigger volume (10l) which resulted in less increase

After attempts to characterise the sensor output this way, the results were unsatisfactory and it was evident that more expensive apparatus, such as an environmental chamber, would have to be used

Difficulties were also encountered while characterising the gas sensor output using the environmental chamber This experiment was abandoned and it was decided to use another PIC to simulate the sensors (temperature, humidity and gas sensor) in order to validate the methods (Figure 7 7) Three very precise potentiometers with dials were chosen and connected to the PIC analogue channels (ADC) The idea was to use these inputs to simulate the sensor output when affected by the temperature and humidity The calculated sensor output was then converted into digital form using PWM on the PIC and a 10 bit digital to analogue converter (DAC) with a duty cycle that corresponds to the analogue output (voltage)

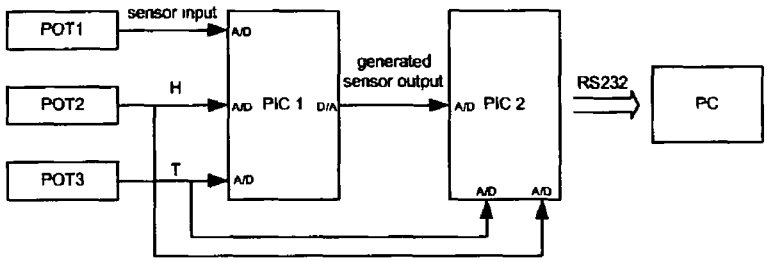

Figure 7 7 Experiment setup

Surveying the range of Figaro gas sensors it was discovered that their output characteristics are typically described by the equation x^k where k varies between 1 8 and 2 5 depending on the sensor $k = 2$ is chosen as it is easy to implement it on the PIC microcontroller and find the inverse transfer function with a *sqrt* math function built in C, thus verifying the end result of the linearising calibration The equation used has a form

$$f(x, H, T) = x^2 + a(T - T_{ref}) + b(H - H_{ref}) \tag{7 1}$$

where a and b are compensation coefficients and T_{ref} and H_{ref} are the values of temperature and humidity for which the sensor was calibrated These coefficients are then calculated on the second PIC (used to perform the compensation and linearising calibration) together with the linearising calibration coefficients and compared with the set values

This alternative way of validation also resulted in reducing the C code on the second PIC used by SHT11 and the gas sensor and enabled testing the Gauss

elimination method and PPC (in its non-recursive form as the PIC cannot handle recursive function calls) with 3 calibration coefficients that gave the errors less than 1% These errors represent not just the errors of the linearising calibration and PPC method and compensation but also those of the 10 bit ADC, the 10 bit DAC and other conversions that had to be performed on the simulating PIC as well There was not enough memory for the testing PPC for 4 or more calibration coefficients

The experiment yielded good results and validated the effectiveness of the proposed module Calibration and compensation coefficients calculated on PIC2 matched the set coefficients on PIC1, with negligible errors in calculation

Chapter 8

Conclusions and Recommendations

8.1 Conclusions

In this thesis various mathematical methods for the linearisation of sensor transfer functions were presented The Gauss elimination and progressive polynomial calibration were chosen for the design of the linearisation and calibration module that would be implemented as part of the generic embedded sensor interface (GESI) The other module that was designed and implemented was the compensation module Also proposed generic embedded sensor interfacing requirements are given

In the opening chapters, a short summary of sensor characterisation was presented, their interfacing requirements and identified the market demands The growing demand for applications using gas sensors, environmental sensors and SAW sensors was recognised

Based on the sensors' interface requirements it was decided that the generic embedded sensor interface (GESI) should facilitate the following excitation of the sensor, amplification, minimum 10 bit ADC, temperature and humidity compensation, calibration, response linearisation, self test, digital signal processing, analogue current loop (4-20mA) interface, serial digital interface e g RS232, RS48 and IR communication

Semiconductor gas sensors with linearisation, compensation for temperature and humidity, analogue current loop link, and digital serial link were chosen due to their highly non-linear characteristic and need for compensation thus having the greatest potential for testing the linearisation, calibration and compensation module discussed in the thesis

The chosen semiconductor gas sensor was somewhat taken as a "black box" to test different linearisation methods curve fitting (with Taylor polynomial interpolation, Gaussian elimination, Lagrange interpolation) and the progressive polynomial calibration (PPC) method These linearisation techniques were modelled and simulated using MATLAB/SIMULINK Modelling the Taylor polynomial interpolation in MATLAB yielded the conclusion that MATLAB's Symbolic Toolbox was not suitable for modelling since it is not based on numerical methods that would be used on the microcontroller (as described in Chapter 5 of the thesis)

Taylor polynomial interpolation was abandoned because it requires equally spaced points, which in practice is hardly possible to achieve, and even a small digression causes a large error Gaussian elimination, Lagrange interpolation and PPC are more flexible However, they require some knowledge of the sensor transfer function and adjusting the measurements accordingly during the calibration process which can be a disadvantage since the sensor transfer function is usually unknown They also have the same drawbacks selection of the calibration measurements in order to get the best approximation and avoid the global dependence on a local properties, and polynomial oscillation, characteristic of the Lagrange interpolation and the PPC Adding extra calibration points after a certain point is to no avail as it results in a more complicated model and even larger polynomial oscillations However, the progressive polynomial method demonstrated good results for certain functions

Among the numerous mathematical methods available, Gaussian elimination and PPC gave the best linearisation results and were deemed the most suitable for the design of the linearisation and calibration module This module was then implemented in C as it would be included in the GESI The use of the C code is

justified by its portability whereas assembler varies from processor to processor

In addition to linearisation and calibration, compensation of the sensor output is often required The digital approach was chosen over analogue compensation techniques as the latter adds electronics to the compensation module and it is inconvenient for compensation in more than one dimension The compensation module presented performs the real-time compensation before the linearisation and any other digital processing, thus the recalibration due to different ambient conditions is not necessary The method was validated by the experimental results in Chapter 7 and also implemented in C on the microcontroller as the compensation module of the GESI

Nevertheless, the chosen TG832 gas semiconductor sensor was found not to be suitable for testing as it is not as selective as suggested by the data sheet, and it is extremely susceptible to any gas present, responding to a wide range of gases In general, the sensor output was higher than expected and was increasing over time with the temperature and humidity held constant Moreover, difficulties were encountered while controlling the ambient conditions, i e humidity and temperature For testing and validation of the proposed calibration and compensation methods it was mandatory to vary only one variable at a time while keeping the others constant As humidity is very dependent on temperature and the gas concentration is dependent on humidity and temperature it was difficult to control the gas level, humidity and temperature at the same time without utilising expensive equipment such as an environmental chamber Encountering similar problems even with the environmental chamber, another PIC was used to simulate the sensors and validate the proposed methods Methods are validated but their realisation would require the code optimisation as memory problems were evident

8.2 Recommendations

The following recommendations are based on the simulation and experimental results

1 Since none of the linearisation methods is good enough for all types of functions it would be desirable to have two or more linearisation methods implemented on the microcontroller or the DSP chip Alternative methods can be then selected through the GUI, as well as the number of the calibration points, by the person performing the calibration process Such a person would need some knowledge of the sensor output in order to choose the appropriate method

2 The progressive polynomial calibration method is described with two recursive functions *calling* each other However, simple microcontrollers provide no support for recursive function calls A stack emulation is possible but it implies programming using assembler or using a compiler that has built-in stack emulation Both cases result in higher memory consumption, which is unacceptable as the memory resources are very restricted It is possible to use more powerful DSP chips that provide multiple recursive calls The other possibility is, if restricted memory resources permit it, to encode every calibration function as a separate function

3 Based on the recommendations above and the fact that the GESI will have to have more than a 10 bit ADC to be able to handle high speed signals, implementation of the calibration, linearisation and compensation module on a DSP chip should be considered to reduce the end cost of the electronics

4 As the compensation module is part of the GESI, it is necessary to bear in mind that there are different sensors, some of them do not require compensation of the output while others can exhibit linear or non-linear dependence on the compensated parameter That dependence can be well described with the second order polynomial (3 coefficients) and there is no need to further expand the model, making it more complex To simplify the model,

one of the curve fitting techniques used for linearisation can be used to find those three coefficients rather than writing *new* functions, hence saving the memory resources The second order polynomial can suit all sensors well and provide good compensation for both linear and non-linear variations in the sensor output In addition, the use of a GUI for selecting the number of the compensation coefficients could be considered However, where the behaviour of the sensor output with the compensated parameters is not well known, it is "safer" to use a maximal number of compensation coefficients available within the module (i e three)

5 As a part of the generic embedded sensor interface, the next logical step would be to test the program for calibration, linearisation and compensation on different sensors and validate the design of the modules

Bibliography

[1] P D Wilson, S P Hopkins, R S Spraggs, I Lewis, V Skarda, and J Goodey Applications of a universal sensor interface chip (USIC) for intelligent sensor applications In *IEE Colloquium on Advances in Sensor Applications of a Universal Sensor Interface*, 1995

[2] Texas Instruments DSP Village Homepage `http //dspvillage ti com`

[3] N B Jones Exciting times for dsp *Sensor Review*, 17(1) 13–20, 1997

[4] Amra Pašić Survey of present sensor measurement technologies Technical report, PEI Technologies, 2001

[5] H Kiesele and M H Wittich Electrochemical gas sensors for use under extreme climatic conditions *Drger Review 85*, April 2000

[6] W E Bulst, G Fischerauer, and L Reindl State of the art in wireless sensing with surface acoustic waves *IEEE Transactions on Industrial Electronics*, 48(2), 2001

[7] `http //www weidmuller ca/interface/stotries/1_97/loop htm`

[8] Rs232 data interface a tutorial on data interface and cables `http //www arcelect com/rs232 htm`

[9] `http //www jbmelectronics com/signal htm`

[10] Quick reference for rs485, rs422, rs232 and rs423 Homepage RS485 com `http //www rs485 com/rs485spec html`

[11] Jan Stanek Introduction to rs 422 & rs 485 Hardware Server Homepage `http //www hw cz/english/docs/rs485/rs485 html`

[12] Les A Kane Fieldbus finally arrives *Hydrocarbon Processing, Control Editorial*, 77(11), 1998

[13] Foundation fieldbus wiring tutorial Relcom Inc Homepage `http //www relcominc com/fieldbus/tutorial htm`

[14] `http //www cwt vt edu/wireless_faq/default htm`

[15] `http //archives e-insite net/archives/ednmag/reg/1996/060696/ 12df1 htm`

[16] Manu Sharma Powerhne communication? Shocking? it isnt! *Voice&Data*, March, 2001

[17] Introduction to radio frequency and microwave concepts and applications `http //vig pearsoned com/samplechapter/0130279587 pdf`

[18] Jon Wu Hyung Kyung Song, Ruchi Tyagi Infrared wireless `http //mama indstate edu/users/tyegir/`

[19] Department of Electrical and Computer Engineering Homepage, Worcester Polytechnic Institute `http //www ece wpi edu/courses/ee535/hwk97/ hwk4cd97/husain/ook html`

[20] Jaap Haartsen Bluetooth-the universal radio interface for ad hoc, wireless connectivity *Ericsson Review*, 1998

[21] Bluetooth Technology Net `http //www blue-tooth-technology net/ profile html`

[22] Bluetooth health statement Ericsson Homepage `http //www ericsson com/health/statement/Bluetooth%20health%20statement pdf`

[23] `http //www blue2space com/bluetooth html`

[24] The fiber guide - a learning tool for fiber optic technology Communication Specialties Inc Homepage `http //www commspecial com/fiberguide htm`

[25] Sergio Franco *Design With Operational Amplifiers and Analog Intergrated Circuits* McGraw Hill, third edition, 2002

[26] Ron Mancini *Op Amps For Everyone Design Reference* Von Hoffmann Graphics, 2000

[27] Gert van der Horn and Johan L Huijsing *Integrated Smart Sensors Design and Calibration* Kluwer Academic Publishers, 1998

[28] Jack G Ganssle Self calibrating systems *Embedded Systems Programming*, 1990

[29] David W Brooks Model systems - calibration (student notes) Center for Curriculum and Instruction, University of Nebraska-Lincoln `http //www cci unl edu/Chemistry/hbb/HBBPactCD/NModel htm`

[30] Feedback isolation augments power-supply safety and performance (appl notes) Maxim Dallas Semiconductor Homepage `http //www maxim-ic com/appnotes cfm/appnote_number/664`

[31] William P Klein Applications of signal isolation *Sensor Magazine*, 17(4), 2000

[32] Analog signal isolation EDN Homepage `http //www edn com/archives/ 1998/060498/12df_01 pdf`

[33] `http //www aps anl gov/icalepcs97/paper97/p192 pdf`

[34] `http //www asott nsw edu au/Gear_intrins htm`

[35] `http //www compliance-club com/archive1/000622 html`

[36] Theodore J Rivlin *An Introduction to the Approximation of Functions* Dover Publications, Inc , New York, 1981

[37] S L Salas and Einar Hille *Calculus One and Several Variables Part II* John Wiley & Sons, Inc , 1978

[38] Melvin J Maron and Robert J Lopez *Numerical Analysis a Practical Approach* Wadsworth Publishing, California, USA, 1991

[39] John H Mathews Module for lagrange polynomials `http //math fullerton edu/mathews/n2003/LagrangePolyMod html`

[40] Carl de Boor *A Practical Guide to Splines* Springer-Verlag, New York, 1978

[41] John R Rice *Numerical Methods, Software, and Analysis* McGraw-Hill, New York, 1983

[42] Jerome Hahn Taylor approximation and m-files `http //bradley bradley edu/~jhahn/Note3 pdf`

[43] Erwin Kreyszig *Advanced Engineering Mathematics* John Wiley & Sons, Inc , 1993

[44] L V Atkinson and P J Harley *An Introduction to Numerical Methods with Pascal* Addison-Wesley, 1983

[45] Microchip PIC16F87X data sheet Microchip Technology Inc

[46] MAX220-MAX249 +5V-powered, multichannel RS-232 drivers/receivers MAXIM

[47] Norbert Kockler *Numerical Methods and Scientific Computing Using Software Libraries for Problem Solving* Oxford Science Publications, 1994

[48] Relative humidity sensor hs1100/hs1101 datasheet Humirel Homepage `http //www humirel com/product/fichier/HS1101-HS1100 pdf`

[49] Tgs 832 datasheet Figaro Sensor Homepage `http //www figarosensor com/products/832pdf pdf`

[50] A(n incomplete) catalog of sensor types Custom Sensor Solutions Inc Homepage `http //www customsensorsolutions com/senstype htm`

[51] Mgsm3000 datasheet Microsens Homepage `http //www microsens ch/ products/pdf/MGSM3000 pdf`

[52] Dong Hyun Kim, Ji Young Yoon, Hee Chan Park, and Kwang Ho Kim CO_2 - sensing characteristics of SnO_2 thick film by coating lanthanum oxide *Sensors and Actuators B*, (62), 2000

[53] Gabor Harsanyi Highly sensitive and selective ammonia sensor for environmental monitoring and workplace safety applications Technical University of Budapest, Department of Electronics Technology http //www ett bme hu/research/gassens

[54] http //ion ipc uni-tuebingen de/weimar/research/main_topics/gassensors/transducer html

[55] International Sensor Technology Homepage http //www intlsensor com

[56] Microsens integrated sensor selection guide Microsens Homepage http //www microsens ch/products/gas htm

[57] http //ds4 koreasme com/viewproduct_1_e html

[58] http //www product-search co uk/iet-pub com/features/jan2000/isc shtml

[59] Infrared technology for detecting combustible gases Delphian Corporation Homepage http //www delphian com/infrared%20sensors htm

[60] http //optoelectronics perkinelmer com/Downloads/Gas_detection pdf

[61] Types of gas sensors R&C Instrumentation Homepage http //www prei co za/istsensdisc html

[62] http //www ertresponse com/sops/2114 pdf

[63] http //www oico com/descpid htm

[64] B Hok, A Bluckert, and J Lofving Acoustic gas sensor with ppm resolution In *In Proceedings of the 13th European Conference on Solid-State Transducers (EUROSENSORS XIII)*, 1999

[65] A M Ferber, P Ohlckersn, H Rogne, and M H Lloyd A miniature silicon photoacoustic detector for gas monitoring applications In *MTEC 2001 International Conference on Sensors & Transducers, Birmingham*, 2001

[66] Ralph W Bernstein, Alain Ferber, Liv Furuberg, and Christian Boccaccio Development of a miniature silicon photoacoustic gas sensor In *Proc 2nd Congress and Exhibition for Optical Sensor Technology, Measuring Techniques, Electronics (OPTO'96)*, 1999

[67] http //www oslo sintef no/ecy/7230/environment shtml

[68] Photoacoustic gas sensing silicon microsystems Homepage of 54 7 http //www fifty-four no

[69] http //www sintef no

[70] Dana Romero Humidity Wolfram Research Homepage http // scienceworld wolfram com/physics/Humidity html

[71] Honeywell Homepage http //content honeywell com/building/components/Hycal_Html/Ppt%20view/hc_trng1

[72] Humidity sensors for industrial applications The International Instrumentation and Control Engineering Website http //www iceweb com au/Analyzer/humidity_sensors html

[73] Relative humidity sensors for electronic data loggers The National Museum of Denmark Homepage http //www natmus dk/cons/tp/datalog/datlog4 htm

[74] Denes K Roveti Choosing a humidity sensor A review of three technologies *Sensor Magazine*, 8(7), 2001

[75] Denes K Roveti Humidity sensor A review of three technologies Sensors Online Magazine http //www sensorsmag com/articles/0701/54/index htm

[76] Hygrometer Principle of operation chilled mirror hygrometers Yankee Environmental Systems Inc Homepage http //www yesinc com/products/data/cmh/index html

[77] Jerri Jefferies Product quality improvement with correct moisture measurement in thermal processes using electrolytic hygrometers MEECO Inc Homepage http //www meeco com/pages/a_art5 htm

[78] Julian W Gardner *Microsensors Principles and Applications* John Wiley $ Sons, 1994

[79] Thermocouples overview SensorPulse Corp Homepage http //www sensorpulse com/ProductsOld/sensors/Thermocouples_Overview html

[80] Irwin Bluestein Understanding contact temperature sensors *Sensor Magazine*, 6(1), 1999

[81] Ic temperature sensors find the hot spots (appl notes) Maxim Dallas Semiconductor Homepage http //www maxim-ic com/appnotes cfm/ appnote_number/689/ln/en

[82] G C M Meijer Thermal sensors based on transistors *Sensors and Actuators A*, (Vol 10), 1986

[83] The Savannah River National Laboratory (SRNL) Homepage http // www srs gov/general/scitech/srtc/srtchtm/xtechsol htm

[84] Artur Dybko Fiber optic chemical sensors Chemical Sensors Research Group, Warsaw University of Technology http //www ch pw edu pl/ ~dybko/papers/opto/paper97_3 htm

[85] Ivan Melnyk, Derek Montgomery, and Doug Short Transformer protection *Electrical Business magazine*, December 2003

[86] Per Eklund and Staffan Rydblom Fiber optic sensors Technical report

[87] Temperatures com Inc Homepage http //www temperatures com

[88] Acoustic method for measuring room temperature distribution Takenaka Corporation Homepage http //www takenaka co jp/takenaka_e/news_ e/pr0101/m0101_01 htm

[89] What is orp? ENVIROEQUIP Homepage http //www enviroequip com/quipnotes/ORP htm

[90] Mike Ross Treating water at orp speed Sensorex Homepage http //www sensorex com/support/technical_articles/ treating_water_at_orp_speed html

[91] N Lallemant, F Breussm, and R Weber The oxy-natural gas combustion page IFRF NET http //www ifrf net/oxyflam

[92] Technical articles Automated Aquarium Systems Inc Homepage http //www automatedaquariums com/aastech htm

[93] The beginners guide to ph pH-measurement co uk Homepage http // www electrochemistry co uk/pH_guide_p1 htm

[94] Mike Ross ph electrode performance Sensorex Homepage http //www sensorex com/support/technical_articles/ pH_electrode_performance html

[95] 1996 1999 Sensorex Operating principle Flat surface ph electrodes Pathfinder Instruments Homepage http //www pfinst com/ operprinciplesflatelectr htm

[96] http //www ch pw edu pl/~dybko/csrg/fiber/operating html

[97] http //www pme com/fcsensor htm

[98] Orion conductivity theory Thermo Electron Corporation Homepage http //www thermo com/eThermo/CMA/PDFs/Articles/ articlesFile_11377 pdf

[99] http //www danfoss com/Analytical/pages/oxy%20principle htm

[100] Bill Drafts Acoustic wave technology sensors Sensors Online Magazine http //www sensorsmag com/articles/1000/68/main shtml

Appendix A

Sensor Characterisation Survey

A.1 Gas sensors

Gas measurement is extremely important in a wide range of applications: food industry, Heating, Ventilation, and Air Conditioning (HVAC), air pollution, water pollutant gases, gas measurement in hazardous area, medicine (especially measurement of arterial blood gases (ABG) which has an important role in the process of clinical evaluation), process control, petrochemical and refinery industries. The most common gas sensors are semiconductor types (due to their good characteristics), pellistor types for measuring combustible gases and lately infra red (IR) gas sensors as they are now cheaper, more sensitive, more accurate and less power consumptive. IR sensors are mainly used in process control, environmental, HVAC, laboratory and medical applications, especially when high precision is needed. Also monitoring for automotive emissions checks is becoming standard application for IR gas sensors. The market for thermal conductivity gas sensors is very small because many sold 20 years ago are still working fine. They rarely fall out of calibration, need very little maintenance and there are no chemicals to denature or optics to misalign. In addition to these most common types, a wide range of different gas sensors was investigated.

A 1.1 Electrochemical type

This sensor will respond to gases that can be electrolytically reduced or oxidized on a metallic catalyst such as platinum or gold (Figure A 1) [5]. The gas being measured diffuses through a hydrophobic, gas permeable membrane into the electrolyte of the sensor. In the electrolyte enclosure there is a measurement electrode, a counter electrode and a reference electrode. A potentiometer ensures

that there is a constant voltage (working potential) between the measurement electrode and the reference electrode. The counter electrode enables a current to flow through the sensor cell.

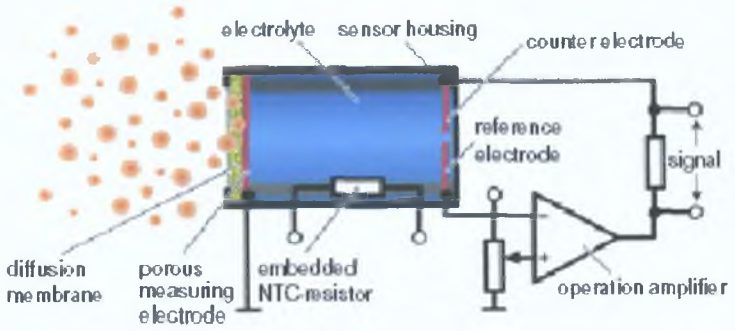

Figure A.1: Electrochemical gas sensor.

Typical gases measured are O_2, CO, NO_2, NO, and H_2S, and organic vapors such as alcohols, aldehydes, or ketones [50]. Typical sensitivities are in the 3-30 ppm (part per million) range, but some proprietary sensors are capable of detecting as little as 2 ppb (part per billion) of gases such as ozone, NO_2 or H_2S. Selectivity is generally modest unless auxiliary methods are used [50]. A selective sensor will respond only to the gas that is set out to measure; it will not respond to other gases that it is likely to encounter. Gas sensors in general respond to a range of gases but this may be enhanced by using auxiliary methods that eliminate or avoid interfering gases. Typical properties are:

- Some sensors require a bias voltage of 0-0.5V on an electrode.
- Sensor output current is proportional to gas concentration; current amplitude is in the μA range; current output is linear.
- Response time is in seconds (s).
- Interface must have low noise, low stray current, high gain.
- Output is effectively dc and interface speed is not an issue.
- Low power consumption.
- Not affected by humidity.

- Suitable for low concentration ppm ranges

A.1 2 Semiconductor (solid state) resistive type

Used for measuring O_2, CO, NO_x, H_2, hydrocarbons (i e CH_4), solvents The sensitive element is a resistive metal oxide layer deposited on an integrated resistive heater [51] The detection of specific gases is based on a conductivity change of the sensing element at an appropriate temperature using a dedicated mode of operation A semiconductor slug is held at high temperature and its resistance changes in response to gas concentration The sensor usually consists of a heater element and a semiconductor slug There is a serious problem with baseline drift (gradual change in the sensor electrical signal, in standard conditions, when no gas is present) [50] A typical semiconductor sensor is shown in the Figure A 2 [52]

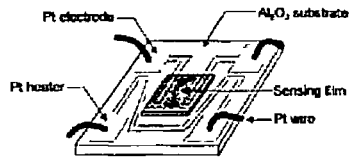

Figure A 2 Thick-film semiconductor sensor

Typical sensor types are[53]

- Thin-film SnO_x resistor elements sputtered on mica substrate
  - Typical operation temperatures are 100^0C-500^0C
  - The sensor signal decreases significantly with increasing operation temperature Humidity hardly affects the signal when detecting NO_2
- Thick-film SnO_2 resistor elements prepared from organo-metallic pastes on alumina substrates[53, 54]
  - Typical operation temperatures for SnO_2 based sensors are 200^0C-400^0C
  - The sensor signal decreases significantly with increasing operation temperature Humidity hardly affects the signal when detecting NO_2

- Electroactive conducting polymers (ECPs) deposited electrochemically and chemically on thick-film substrates, operated at room temperatures

The heater has important effects on semiconductor gas sensors because the working temperature is relatively high, and its distribution on the substrate surface must provide equal temperature distribution for the overall sensor and reference resistor film Sensor resistance is inversely proportional to gas concentration The relationship is non-linear and follows an inverse square law approximately The sensor is very sensitive to ambient temperature and humidity

Typical properties are

- Interface must convert resistance in 1kΩ to 20kΩ range to voltage, invert, linearise relationship and compensate for temperature and humidity
- Output is effectively dc and speed is not an issue
- Response time is less than 30s
- High sensitivity at low concentrations, lower detection limit less than 5 ppm
- Preamplifier and low pass filter are needed

A.1.3 Catalytic (pellistor) type

This type is normally used for detection of combustible gases such as iso-butane or methane They detect combustible gases in higher concentrations (above 1000 ppm) only They are not selective and will respond to a wide range of combustible gases However, since they only detect higher gas concentrations, they are not prone to interference from many toxic gases[55]

A catalytic sensor consists of a sensor element and compensation element When a combustible gas comes into contact with the gas sensor element, the gas is oxidized The oxidation reaction generates heat and the temperature of the platinum coil, which is a sensor element (Figure A 3), rises This temperature rise is directly

proportional to the concentration of the gas. The gas concentration can be measured by using a bridge circuit, measuring the difference in potential between the sensor element (detector) and the compensation element (compensator)[56].The use of a pair of matched elements (detector and compensator) cancels out changes due to ambient temperature, pressure and humidity.

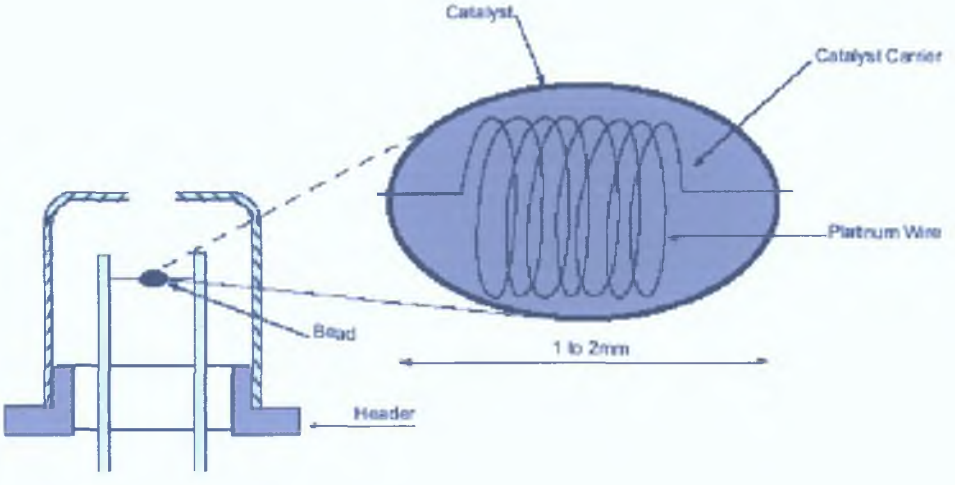

Figure A.3: Cross section of a catalytic element

Characteristics[57]:

- Response time less than 10s.
- At low gas concentrations, output is a linear function of gas concentration.
- Sensor output is in mV range.
- Interface circuitry must amplify low level signals.

A.1.4 Thermal conductivity type

Used for measuring concentration of combustible gases, general gases, methane, natural gas, inert gases.

The thermal conductivity sensor (Figure A.4) measures the concentration of a specific gas between a hot surface resistor (R_m) and an ambient temperature ref-

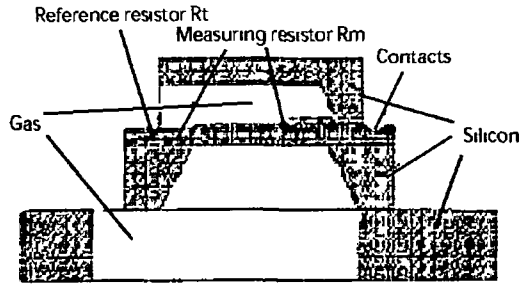

Figure A 4 Cross section of a thermal conductivity gas sensor

erence resistor (R_t) using the thermal conductivity coefficient of the gas itself[56] The sensor structure consists of an integrated heater located on a thin electrical and thermal insulating membrane

A variation of this type of sensor is similar to the catalytic type and consists of two elements, both comprised of a wire coil (in some cases the coil is coated) One element (detector) is exposed to the atmosphere, whereas the other element (reference) is sealed in a standard gas atmosphere such as nitrogen The elements are heated to an operating temperature of approximately 250^0C Heat is transferred from the element to the surrounding gas The amount of heat transferred depends on the thermal conductivity of the gas When the heat transfer occurs, the element's resistance changes and that change is measured using a bridge circuit[58]

- Very stable long term operation (low long-term zero and sensitivity drift)
- Low power consumption (typical heating power less than 6 mW)
- Response time 20s
- Temperature and humidity compensation are required

A.1 5 Infrared (IR) type

This type is based on the principle that gases absorb light energy at a specific wavelength, typically in the infrared range corresponding to the resonance of the molecular bonding between dissimilar atoms Gases which contain more than one type of atom absorb IR radiation [59] When infrared light strikes a substance, the radiation is transmitted, reflected or absorbed in varying degrees, depending upon the substance and the wavelength of the radiation Since radiation is a form of energy, when a molecule is excited by light, the energy of the light is absorbed by the molecule The absorbed energy causes an increased level of vibration or rotation Each molecule can vibrate and rotate in certain patterns, and for each pattern there is an associated amount of energy of motion A molecule can only absorb energy from a photon if the energy matches precisely to an "energy state" of that molecule Therefore gases such as carbon dioxide, carbon monoxide, methane, sulphur dioxide, nitric oxides, can be detected by this means but gases such as oxygen, hydrogen and chlorine cannot because they oscillate in high energy states They cannot absorb infrared radiation IR sensors are mostly used for carbon dioxide, because no other sensing method works reliably for this gas

Different gases absorb infra-red light at different frequencies The sensor in Figure A 5 typically consists of a light source such as an incandescent lamp, a chamber for containing a sample gas and a detector (DET) with an optical filter [58] Light is passed through the chamber and impinges on the detector As the sample gas is introduced into the chamber, it absorbs the light at a specific wavelength or range of wavelengths The filter is a non-dispersive optical component, and allows the sensor to identify the gas based on the unique fingerprint in the absorption spectrum The narrower the bandwidth of the filter, the smaller the wavelength range and therefore the greater the specificity of the sensor The amount of absorbed light is related to the gas concentration and follows Beer's Law

The light emitted into the gas detection section (DET) varies with the age of the light source, the electrical supply, and various other influences Most designs have

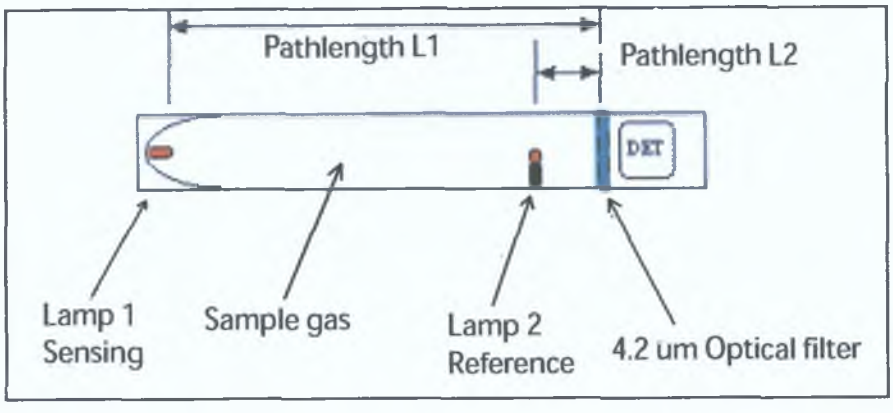

Figure A.5: Infrared sensor

a reference channel to monitor system integrity. This reference is a second source which verifies the strength of the full light signal from the source [59]. Also, when the temperature varies, the optical thickness of the filter layers change due to thermal expansion and due to the changes in the refractive indices [60, 59]. Water vapour is transparent to infrared from 3 to 4.6 μm but shows significant absorption outside this band in the mid-infrared range. This absorption appears like gas to the detector and can cause false readings. In addition, water vapour can condense on the optics or in the path and cause the beam to be deflected or diffracted [59]. Infrared systems are also affected by changes in pressure. As pressure increases more molecules are packed into the path and therefore more infrared radiation is absorbed. [59].

A.1.6 UV open path type

Photoionization detectors (PIDs) offer fast response for many Volatile Organic Compounds (VOC's), including, among others, benzene, vinyl chloride, phosphine, toluene, xylene, and hexane.

PIDs operate by ionizing the target gas with ultraviolet (UV) radiation, and then collecting the ions across a high voltage plate which produces an electrical signal proportional to the gas present(Figure A.6) [61]. The optical path can be up to 100m. Temperature compensation is needed[61].

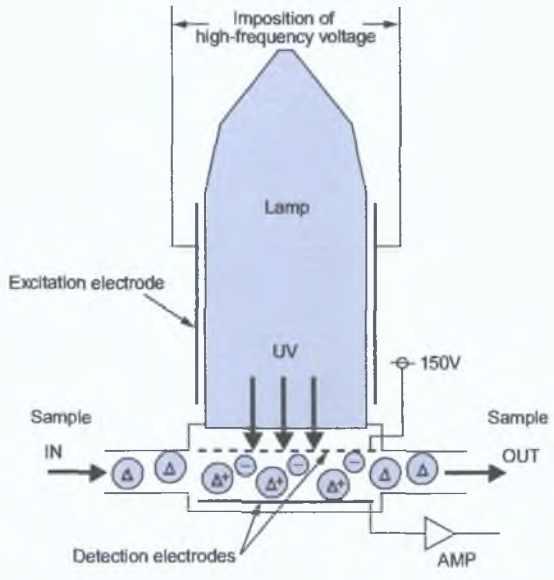

Figure A.6: Operating principle of Photoionization detector (PID).

The analyzer will respond to most vapours that have an ionization potential less than or equal to that supplied by the ionization source, the UV lamp [62]. Photoionization occurs when an atom or molecule absorbs a photon of sufficient energy to release an electron and form a positive ion. This will occur when the ionization potential (measured in eV) of the molecule is less than the energy of the photon.

The sample stream flows through the detector's reaction chamber where it is continuously irradiated with high energy ultraviolet light [63]. When compounds are present that have a lower ionization potential than that of the irradiation energy (10.2 eV with standard lamp), they are ionized. The ions formed are collected in an electrical field, producing an ion current that is proportional to compound concentration.

A.1.7 Acoustic type

The speed of sound changes in different gases The velocity of sound c in gases is given by $c = (RT\gamma/M)^{1/2}$[64], where R is gas constant, T is absolute temperature, M is the molecular mass of the gas and γ is ratio between the heat capacities at constant pressure and volume The speed of sound in a gas, along with the tube length, determines the resonance frequency of a tube (Figure A 7) The resonance frequency is the frequency at which the gas in the tube will most vigorously vibrate when driven by an external source A pair of ultrasonic transmitter/receivers operating at 40kHz are mounted at the ends of a cylindrical cell, the mantle of which is enclosed by a heating coil and an inner and outer mantle (not shown in the Figure) of copper and a thermally insulating material, respectively A control loop, including a thermistor, maintains the cell at approximately 40^0C Passive exchange of gas to the sensor cell is provided by openings in the mantle

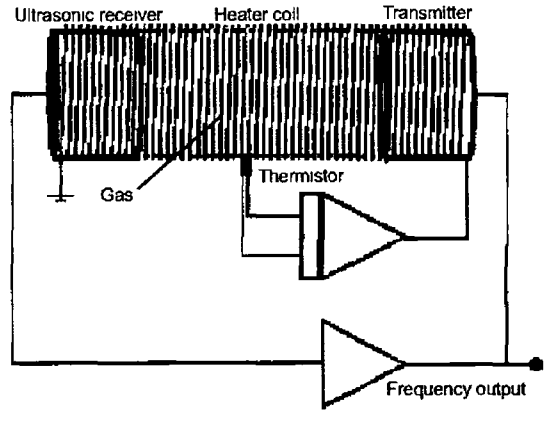

Figure A 7 Basic design of a high resolution acoustic sensor

The effect of temperature on the speed of sound is compensated electronically Humidity compensation may be performed with an appropriate humidity sensor

The sensor's linear dimensions are approximately related to the operating frequency, whereas response time is related to volume The response time can thus be expected to decrease dramatically upon miniaturization[64]

A.1.8 Photoacoustic type

Many gases are measured using this technique, such as CO_2, N_2O, CH_4, SF_6, R404a and R134a (refrigeration gases) [65].

As described earlier, when a gas is exposed to infrared (IR) light, it absorbs radiation within its own characteristic absorption spectrum [66]. Each gas has a unique IR spectrum, and strong absorption takes place only at certain wavelengths. The presence of a target gas in the absorption path reduces the light intensity (pressure signal) in a sealed cell [67]. The target gas itself functions as a filter. Transmitted light enters the sealed absorption cell (Figure A.8), which is filled with a target gas [66]. Most of the light corresponding to the gas specific absorption wavelengths gets absorbed. The absorbed radiation is quickly released by relaxation of translational energy ($\frac{1}{2}mv^2$). The absorbed light energy becomes heat energy and the sudden expansion of the gas generates a pressure, or acoustic wave, which is measured by the pressure sensor (microphone).

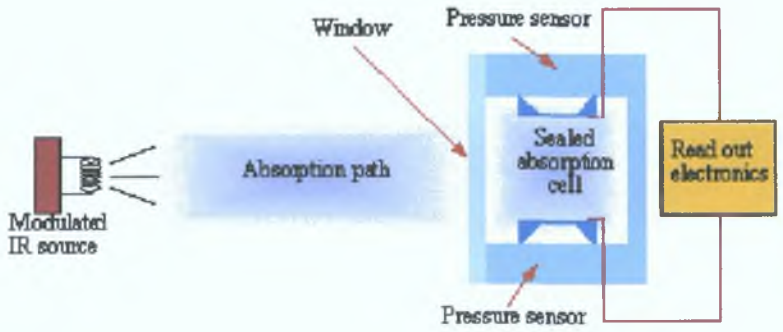

Figure A.8: The principle operation of the photoacoustic gas sensor [68].

In a conventional photoacoustic sensor, the gas to be analyzed has to be pumped continuously into and out of the absorption cell that is sealed with mechanical valves during the measurement [66]. Nowadays, photoacoustic gas sensors based on silicon microsystems technology are available, such as one developed by SINTEF [69]. This sensor does not include valves, pumps or filters. This leads to less wear, longer life-time and lower price.

A.1 9 Other types of gas sensors

Other gas sensors, less common m usage, use other operating principles Amongst them, the most common are fiber-optic, chemiluminescence type, microbalances type, conductive polymer, elastomer chemiresistor type [50]

- **Chemiluminescence type** [50] Certain chemical reactions generate light, which can be measured with great sensitivity The most common application of chemiluminescence in gas detection is the measurement of nitric oxide by reaction with ozone

- **Fiber-optic type** [50] A thin glass or plastic fiber is coated with a thin layer of a compound that will absorb the analyte (the chemical being investigated by an analytical method) When light is passed through the fiber and reflects from its inside surface, some of the light energy extends beyond the surface of the fiber This effect is known as the evanescent wave, and its influence is usually no more than a few nm A simple surface coating may absorb organic gases, changing its refractive index, and the amount of light reflected inside the fiber is changed This is detected by a receiver at the other end of the fiber from the light source Other surface coatings may react with the analyte gas and change colour, which will affect the spectrum of the reflected light The advantage is that no input energy is required to start the absorption process

- **Microbalances type** [50] The simplest form of this sensor uses a quartz crystal which is electronically made to vibrate at its natural frequency The crystal is coated with a material that absorbs the analyte gas The mass of the coating increases and slows down the natural rate of vibration of the crystal The resulting frequency shifts can be measured electronically with great sensitivity The basic microbalance has been elaborated into more sophisticated devices such as surface acoustic wave (SAW) and resonator devices

- **Conductive Polymer type** [50] Certain polymers, such as polyanilines and polythiophenes, are electrically conductive The conductivity changes

when certain gases are absorbed by the polymers. The polymers can also be "tuned" to certain compounds by carrying out the polymerization in the presence of the analyte. They are sensitive to temperature and humidity.

- **Elastomer Chemiresistor type** [50]. These measure the very slight physical expansion of a film of an elastomeric material that occurs when it absorbs a gas. The elastomer, silicone rubber for example, contains electrically conductive particles such as carbon. The concentration of particles is adjusted so that there are relatively few conducting paths through the elastomer. Slight expansion of the elastomer causes some of these paths to be broken, and the electrical resistance rapidly increases.

A.2 Humidity Sensors

Humidity plays a part in every industrial process. The extent to which it does may vary but in most cases it is essential to have it at least monitored and in most cases controlled. In general, humidity sensors need temperature compensation as the humidity is very temperature dependent.

Humidity is the amount of water saturated in air (or another gas) [70]. Absolute humidity gives the mass of water vapor per unit volume of air in the atmosphere [70]. The relative humidity (RH) is the ratio of the partial pressure of water which is actually present to the vapor pressure of water at that temperature, expressed in percent [70].

$$RH = \frac{partial\ pressure\ of\ water\ vapor}{saturation\ water\ vapor\ pressure} \times 100\% \qquad (A.1)$$

In general RH sensors have slow response, between 1 and 20 minutes.

The most common applications[71] are:

- HVAC (air handling equipment, clean room systems, indoor air quality, humidifiers/dehumidifiers and humidity controllers, refrigerant gases),

- instrumentation (weather stations, remote monitoring, monitoring devices that need relative humidity (RH) compensation, data loggers and recorders, environmental test chambers),
- medicine (ventilators, compressors, incubators, surgery areas),
- process control,
- agricultural (incubators, animal confinement buildings, greenhouses),
- moisture detection (sealed instrumentation, fiber optic transceiver enclosures),
- battery applications,
- pipeline drying,
- polymer chip drying,
- various laboratory measurements including measurements in highly refined gases used in microelectronic technology, calibration reference

A variety of commercial and technical criteria dictates which measurement technology (sensor type) is used for any particular application

A 2 1 Wet bulb/Dry bulb psychrometer

Psychrometers are typically used to control climatic/environmental chambers A psychrometer (Figure A 9) consists of a pair of matched thermometers, one of which is in a wetted condition In operation, water evaporation cools the wetted thermometer, resulting in a measurable difference between it and the ambient, or dry bulb measurement

The most commonly used of the psychrometer formulas is the Sprung formula [70] The partial pressure of water vapor in air is proportional to the difference in temperature of the wet bulb and dry bulb When this pressure is calculated, for example using a DSP device or microcontroller, RH can be calculated using the

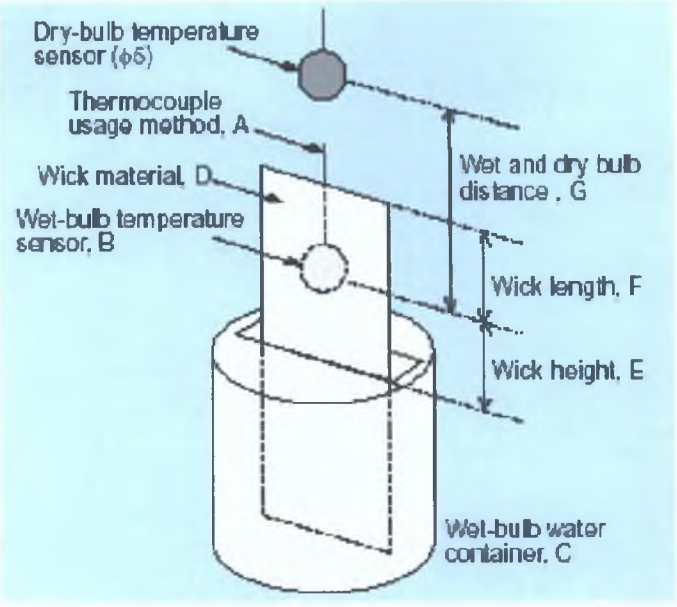

Figure A.9: A Psychrometer

equation A.1 as a ratio of the partial pressure of water vapor and the saturation water vapor pressure at dry-bulb temperature, expressed as a percentage.

Whilst the psychrometer provides high accuracy at near saturation (100% RH) conditions and is simple to use and repair, its accuracy at lower relative humidities (below 20%) is poor and maintenance requirements are intensive. It cannot be used at temperatures below 0^0C and, because the psychrometer is itself a source of moisture, it cannot be used in small, closed volumes.

A.2.2 Bulk polymer resistive sensor type

Polymer resistive sensors are used in HVAC, handheld devices, environmental and meteorological monitoring.

These electrical sensors (Figure A.10) provide a direct, secondary measurement of relative humidity [72]. The quantity that is measured is the simple electrical resistance through, or across the surface of, the polymer, which changes with water content. Changes in relative humidity affect the resistance of the electrode. This sensor also needs an alternating excitation voltage, not for the measurement

but to avoid destroying it by causing one-way electrolytic ion movement in the polymer [73].

Figure A.10: Bulk polymer resistive RH sensor

Bulk polymer resistive sensors are relatively immune to surface contamination. Although surface build up does not affect the accuracy of the sensor, it does have an effect on the response time. Due to the extremely high resistance at RH values of less than 20%, this sensor is generally better suited to the higher RH ranges [72].

The resistive sensor exhibits a non-linear response to changes in humidity[74]. The impedance change is typically an inverse exponential relationship to humidity.

A.2.3 Capacitive type

Polymer capacitive sensors (Figure A.11), similarly to the polymer resistive are used in HVAC, handheld devices, environmental and meteorological monitoring. They can operate over a relatively wide span encompassing high pressure applications. They can also be used to measure moisture in liquids and, due to low power usage, are suitable for use in explosive atmospheres. [72].

Figure A.11: Bulk polymer capacitive RH sensors.

The capacitive RH sensor consists of a thin layer of water absorbent dielectric, that is coated onto a conductive base [73] The layer is then covered with a porous conductive layer As the relative humidity increases the water content of the polymer increases Water has a high dielectric constant This means that the combination of the two electrodes with the water between can store more electric charge This capacitance is measured by applying an alternating voltage across the plates and measuring the current that passes To get appropriate measurements the sensor has to be excited at a frequency between 500Hz and 2kHz

Capacitive sensors have a serious limitation [73] the change in capacitance is small compared with the capacitance of even a few meters of cable This means that the electronic processing has to be completed close to the sensor

This sensor type is ideally suited for use in high temperature environments because it will provide a relatively fast response (around 60s) due to the low temperature coefficient so the polymer dielectric can withstand high temperature At RH values over 85% however, the sensor has a tendency to saturate and become non-linear [73]

The main disadvantage associated with these sensors is that they are secondary measurement devices and must be frequently recalibrated to accommodate ageing effects, hysteresis and contamination [72] Capacitive sensors are also limited by distance, the sensing element can not be located far from the signal conditioning circuitry This is due to the capacitive effect of the connecting cable with respect to the relatively small capacitance changes of the sensor [75]

A.2 4 Saturated salt lithium chloride type

Saturate salt sensors are an attractive proposition when a low cost, rugged, slow responding and moderately accurate sensor is required They are typically used for refrigeration controls, dryers, dehumidifiers, air line monitoring The saturated salt lithium chloride sensor is one of the most widely used dew point sensors Its popularity stems from its simplicity, low cost, durability, and the fact

that it provides a fundamental measurement [72]

The sensor consists of a bobbin covered with an absorbent fabric and a bifilar winding of inert electrodes The bobbin is coated with a dilute solution of lithium chloride An alternating current is passed through the winding and the salt solution causing resistive heating As the bobbin heats, water evaporates from the salt at a rate which is controlled by the vapor pressure of water in the surrounding air As the bobbin begins to dry out, the resistance of the salt solution increases causing less current to flow through the winding and allowing the bobbin to cool This heating and cooling of the bobbin reaches an equilibrium point where it neither takes on nor gives off water This equilibrium temperature is directly proportional to the water vapor pressure or dew point of the surrounding air It is usually measured using a platinum resistance thermometer (PRT, described later) and output directly as a D/F PT (Dew/Frost Point) Dewpoint is the temperature (above 0^0C) at which the water vapor in a gas condenses to liquid water [72] Frost point is the temperature (below 0^0C) at which the vapor crystallises to ice [72] D/F PT is a function of the pressure of the gas but is independent of temperature and is therefore defined as fundamental

If a saturated salt sensor becomes contaminated, it can easily be cleaned and recharged with lithium chloride The limitations of the technology are a relatively slow response time (less than 5 minutes) and a lower limit of the measurement range which is imposed by the nature of the lithium chloride The sensor cannot be used to measure dew points when the vapor pressure of water is below the saturation vapor pressure of lithium chloride, which occurs at about 11% RH [72]

A 2.5 Chilled mirror (optical condensation) hygrometer

Typical applications for the optical condensation hygrometer are medicine, liquid cooled electronics, cooled computers, heat treating furnaces, smelting furnaces, clean room controls, dryers, humidity calibration standards, engine test beds

The actual condensation point of the ambient gas is measured The sensor contains a small metallic mirror, the surface of which is chilled until water condenses

out of the sample gas onto the mirror surface [72].

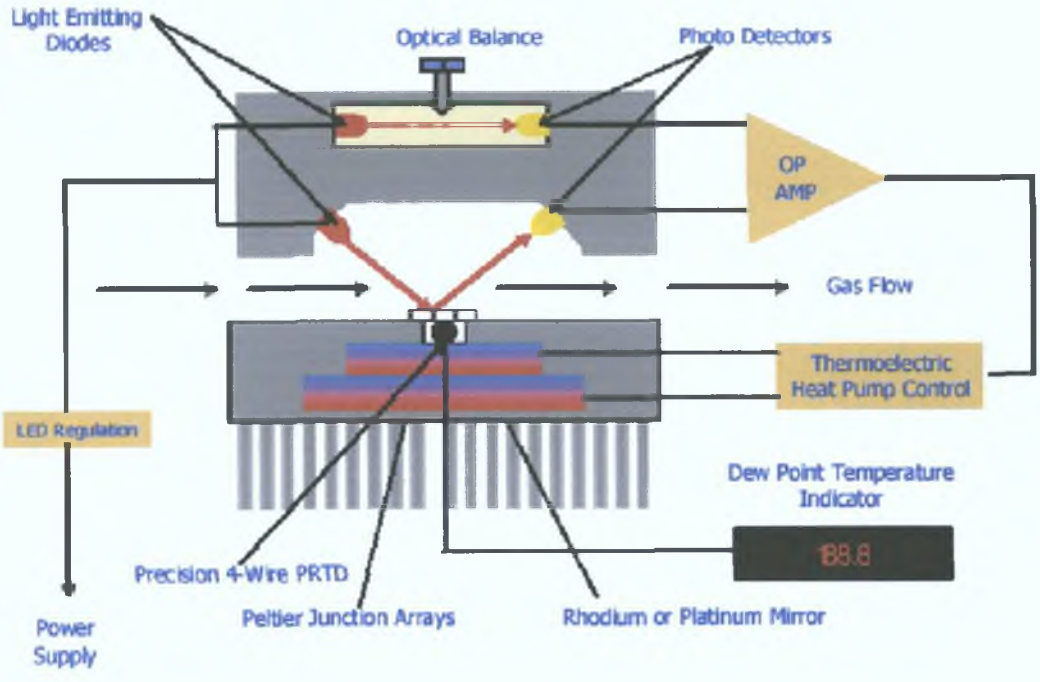

Figure A.12: A schematic of a chilled mirror hygrometer that utilizes optical feedback, thermoelectric cooling and an PRTD.

A mirror is constructed from a material with good thermal conductivity such as silver or copper, and properly plated with an inert metal such as iridium, rubidium, nickel, or gold to prevent tarnishing and oxidation [76]. The mirror is chilled using a thermoelectric cooler until dew just begins to form. A beam of light from a light source (typically from a LED) is aimed at the mirror and the reflection is detected by a photodetector (typically a phototransistor). The schematic is shown in the Figure A.12.

At the occurrence of condensation, the reflected light is scattered and, therefore, reduced at the detector. A control system keeps the temperature of the mirror at the point where a thin film of condensation is maintained. A PRT embedded in the mirror measures its temperature and therefore, the D/F PT temperature. At this temperature, the mass of water on the surface is neither increasing (too cold a surface) nor decreasing (too warm a surface) [72].

Accuracies of $\pm 0.2^0$C are possible with chilled mirror hygrometry. Response times

are fast and operation is relatively drift free [72]

A.2.6 Electrolytic hygrometer

The electrolytic sensor is used in very dry applications and is well suited for use in industrial processes such as ultra pure gas, specialist catalyst, fine chemicals and integrated circuit production Other typical applications for this sensor are ozone generators, dry air lines, nitrogen transfer systems, inert gas welding

Typically, the electrolytic sensor requires that the gas being measured must be clean and should not react with a phosphoric acid solution, although recent developments in the sensor cell technology and sample conditioning systems allow more hostile applications, such as moisture in chlorine to be addressed [72]

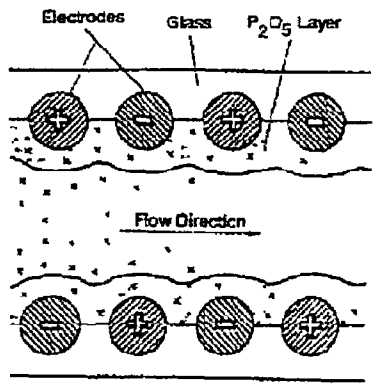

Figure A 13 Operation of electrolytic hygrometer sensing element

The electrolytic sensor (Figure A 13) utilises a cell coated with a thin film of phosphorous pentoxide (P_2O_5), which absorbs water from the gas under measurement The cell consists of a hollow glass tube with two platinum electrodes spirally wound around the inside wall The sample gas is passed through the glass tube and the P_2O_5 film attracts and absorbs all of the moisture When an electrical current is applied to the electrodes, the water vapor absorbed by the P_2O_5 is dissociated into hydrogen and oxygen molecules The amount of current required to dissociate the water is proportional to the number of water molecules present in the sample The accuracy depends on maintaining a controlled and monitored sample flow [77]

A.2 7 Piezo-resonance crystal sensor

The piezo-resonance sensor operates on the principle of RH equilibrium where the absorption of water increases the mass, which directly affects the resonant frequency of the crystal [72]

The sensor has a humidity sensitive coating placed on a resonating crystal surface The crystal resonant frequency changes as the humidity sensitive coating either absorbs or desorbs water vapor in response to changes in the ambient humidity levels This resonant frequency is compared to a similar measurement in a dry gas or to a reference frequency that has been calibrated for % RH [72]

A.3 Temperature Sensors

Temperature measurement, like humidity, is central to almost every industrial process The main areas are process control, HVAC, instrumentation (weather stations, emote monitoring), food service equipment (fryers, grills, toasters, ovens), thermal management (avoid equipment breakdown, extend equipment life), thermal mass flow (gas flow, liquid flow), laboratory (sterilization equipment, environmental chambers, incubators), telecommunications equipment The most common sensors are thermocouples, various types based on the resistance thermometry, semiconductor type (various IC) and fiber optic types

A 3 1 Thermocouples

A thermocouple consists of two different metals and the junctions are held at different temperatures [78] One junction is held at one temperature (sensing), while a second (reference) junction is held at the different temperature (Figure A 14)

A thermoelectric potential ΔV is generated across the junctions The thermoelectric effect is known as the Seebeck effect The open circuit voltage depends upon the difference in the temperature between the sensing and reference point It is related to the Seebeck coefficient or thermopower P_a which depends on the

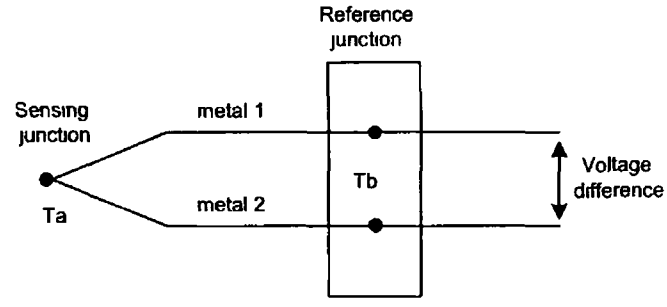

Figure A 14 Thermocouple

material and it is given by

$$P_a = \frac{\Delta V}{\Delta T} \tag{A 2}$$

Thermocouples can be made from a variety of metals and alloys and cover a temperature range of -200^{0}C to 2200^{0}C

The sensing junction (Figure A 14) may be grounded, insulated or exposed Exposed junction thermocouples do not have their junctions protected by a sheath Response time is fastest with this type of thermocouple but the maximum allowable temperature is less than that of a sheathed thermocouple A grounded thermocouple has its sensing junction in contact with a metal surface This is the most common type Where electrical isolation is necessary (measuring in corrosive conditions), an insulated junction is used [79]

They have many disadvantages The wire has much higher resistivity than copper, and despite the fact that a short circuit exists at the hot junction, by the time it reaches the measuring instrument it can become an effective antenna for radiated interference (EMI) [80] The relationship between the thermocouple mV signal and the process temperature is not linear When time response requirements require that the thermocouple junction be grounded, an isolated signal conditioner is essential to prevent the ground loop that would result from a direct connection

A.3.2 Resistance thermometry

A thermistor (Figure A 15) is a thermally sensitive resistor, the resistance of which varies according to its temperature [80] There are thermistors with negative temperature coefficient (NTC) and positive temperature coefficient (PTC) A NTC thermistor is one in which the resistance decreases with an increase in temperature, whereas in a PTC increases with an increase in temperature

Figure A 15 Thermistor

The other type of sensors that change their resistance with the temperature are resistance temperature detectors (RTDs) shown in Figure A 16 They are made by winding small diameter wire on a mandrel and can be made from a variety of metals, the less expensive of which are copper, nickel, and nickel iron alloys [80] Although all these metals have positive temperature coefficients of resistance, each has different characteristics The best choice is platinum (Platinum Resistance Thermometers - PRTs) because it has a relatively high resistivity, is very stable, is inert and nearly linear Their temperature coefficients of resistance differ in magnitude and linearity, and the resistivity of each is different

Figure A 16 RTD

The accuracy of PRT temperature measurement is largely determined by the number of leads used between the probe and the instrument Two leads are often acceptable in the case of short cable runs but three or four leads compensating for

lead resistance variations give the greatest precision The choice of 2, 3 or 4 wires must be made in context with the measuring instrument input arrangement [80]

NTC thermistors have highly non-liner characteristic Like the RTDs, thermistors require external current excitation and significant signal conditioning

A.3.3 Semiconductor circuits

Integrated circuits (IC) temperature sensors are part of every PC and the majority of automotive applications Cellular phones usually include one or more sensors in the battery pack, and notebook computers might have four or more sensors for checking temperatures in the CPU, battery or AC adapter [81] They require no linearisation or reference junction compensation, unlike thermocouples, on the contrary, they often provide reference junction compensation for them They generally provide better noise immunity through higher-level output signals, and some provide logic outputs that can interface directly to digital systems

There are families of IC temperature sensors available today that produce linear current or voltage signals related to temperature over the -40^{0}C to 150^{0}C range [80] More sophisticated temperature sensors include a serial interface, such as the I2C (Inter-Integrated Circuit bus), SPI (Serial Peripheral Interface), or SMBus (System Management Bus), which provides communication with embedded microcontrollers and other digital systems [81]

These sensors employ the principle that the difference in base-emitter voltage V_{BE} of two transistors with different current densities I_s exhibits a linear relation with respect to the absolute temperature [82] If two equal transistors are biased with currents that are scaled by factor of four, as illustrated in Figure A 17, then from equation A 3 and A 4

$$I_{c_{T1}} = I_{s_{T1}} e^{\frac{q}{kT} V_{BE_{T1}}} = 4I_{bias} \tag{A 3}$$

$$I_{c_{T2}} = I_{s_{T2}} e^{\frac{q}{kT} V_{BE_{12}}} = I_{bias} \tag{A 4}$$

follows that the difference in their V_{BE} is proportional to KT/q

$$\Delta V_{BE} = \frac{kT}{q} \ln(4 \frac{I_{s_{T2}}}{I_{sT1}}) \tag{A 5}$$

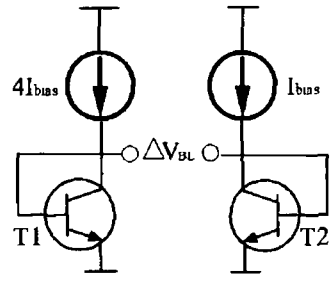

Figure A 17 Simplified functional diagram of integrated circuit temperature transducer

A.3.4 Fiber optic type

The optical-based temperature sensor provides accurate and stable remote measurement of on-line temperatures m hazardous environments and in environments having high ambient electromagnetic fields without the need for calibration of individual probes and sensors Because it does not require either calibration or recalibration, and because its simplicity lends reliability and little need for maintenance, it is well suited to remote or inaccessible locations, such as underseas, geological studies, or space applications [83] Because it is corrosion-resistant, it can be used in food processing, petrochemical processing, nuclear processing plants, paper plants, and the like

The only limits to the use of this system is in environments where temperatures exceed the melting point of the probe, sensor, or fiber optic components Such temperatures might be found in steel or other foundry applications, and in fusion or other high energy research applications [83]

In field application one of the main problems is a power supply for a sensor and its electronics This problem does not occur in fiber optic temperature sensors The measuring probe can be located a long distance from the electronic circuit because the optical signal is guided by fibers and the working principles of the sensor do not require a power supply

There are many types of fiber optical temperature sensors used today, phase modulated and intensity modulated sensors (amplitude modulation of light, e g

sensors based on liquid crystals) or fluorescence decay time. Other sensors utilize changes of dimensions of air bubble, changes in the transmission of semiconductor wafer etc [84].

A fiber-optic temperature sensor, based on fluorescence is shown in Figure A.18 [85]. Pulses of light are transmitted down the fiber, which excites a phosphor material bonded to the fiber tip. The phosphor material glows a different colour. The light is coupled back into the fiber and reflected onto the detector with a colour separation filter. The persistence of the fluorescent afterglow depends on the temperature of the phosphor (generally, the colder it is the longer the afterglow lasts).

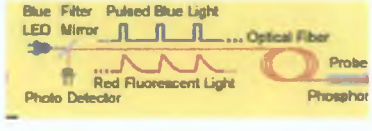

Figure A.18: Fiber optic temperature sensor with based on the Raman effect.

Phase modulated sensors work by comparing the phase of light in the sensing fiber with a reference fiber in an interferometer (measures small displacements of two waves). In general they consist of a coherent laser which injects light into two single mode fibers. If the environment perturbs one or both fibers differently, a phase shift will occur that can easily be detected [86].

A.3.5 Radiation Thermometers (RTs)

Radiation Thermometers (Pyrometers) are non-contact temperature sensors that measure temperature from the amount of thermal electromagnetic radiation received from a spot on the object of measurement [87]. All objects emit thermal electromagnetic radiation and that radiation is characteristic of their temperatures - the hotter an object is, the greater its thermal radiation and the more that radiation shifts toward shorter wavelengths.

A variety of names is given to the devices in this class, e.g.: optical pyrometers, ra-

diation pyrometers, total radiation pyrometers, automatic infrared thermometers, continuous radiation thermometers, line-scanners, thermal imaging radiometers, infraducers, infracouples, fibreoptic thermometers, gold cup pyrometers, surface pyrometers, ratio pyrometers, two-color pyrometers, infra-snakes, or something similar [87]

The radiation thermometer generally consists of an optical system, a detector and a signal conditioner

This group of sensors includes both spot or point measuring devices in addition to line measuring radiation thermometers, which produce one-dimensional (1-D) and, with known relative motion, can produce two-dimensional (2-D) temperature distributions, and thermal imaging, or area measuring, thermometers which measure over an area from which the resulting image can be displayed as a 2-D temperature map of the region viewed [87]

There are numerous types of radiation thermometers but they can be broken down into three broad categories [87]

- Spot measuring thermometers They measure the temperature of a spot at some distance They can also be classified according to their technical type, i e whether they collect and measure thermal radiation in a single spectral region, "single waveband devices", or in two or more wavebands simultaneously, or, they can be classified by the wavebands or center wavelength of the measuring waveband So, too, they can be given distinctive names by their developer, to make them distinctive in some way

- Line measuring thermometers These sensors measure a linear region over a defined angular range If there's an object of sufficient temperature covered by that angle, then the device produces a linear trace along the line "seen" by the sensor Some line measuring thermometers are connected to computers in such a way, that if the object moves perpendicular to the measured line, a series of temperature profiles can be connected together to form a sample of the two-dimensional temperature distribution on the object

- Area measuring thermometers They give a two dimensional temperature map of a surface

A 3 6 Acoustic and Ultrasonic type

The concept is that of measuring temperature in a gas by measuring the speed of sound in that gas In room temperature sound travels at about 340m/s, and this sound speed is usually referred to as Mach 1 [88]

One way of measuring the temperature is to send an ultrasonic pulse down a rod of known expansion and propagation properties By placing slots in the rod at known and calculable distances from the excitation position, one could immerse the rod in a medium of high temperature and then measure that temperature by measuring the reflection times of the pulses from the notches [87]

A.4 Water Quality Sensors

Water quality control employs a large number of sensors amongst which the most common ones are oxidation reduction potential (ORP, or Redox), pH, water conductivity and dissolved oxygen Some of these sensors are described in this section They are used with water treatment in industry (wastewater), agriculture, oceanography, wells and landfills

A.4 1 Oxidation-Reduction Potential (ORP)

Oxidation Reduction Potential (ORP, or Redox) is a measurement of a solution's oxidizing and reducing activity The process of oxidation involves losing electrons while reduction involves gaining electrons ORP is measured in mV [89]

Reduction and oxidation (redox) reactions control the behavior of many chemical compounds in water The life expectancy of bacteria in water is also related to ORP Bacteria and algae essentially are hydrocarbons, and chlorine is a powerful

oxidizing reagent Chlorine destroys bacteria and algae by burning their carbon and hydrocarbon into CO_2 and H_2O [90] Actual ORP levels required for bacteriological control will vary with use of different oxidizers Both concentration and activity of the oxidizer will affect the ORP levels

For example they can be also used in swimming pools, cyanic acid is added to minimize the loss of chlorine The cyanic acid reacts with the hypochlorous acid to bind it in a form that reduces the free available chlorine This chemical binding has the net effect of lowering the concentration of chlorine detected by the ORP electrode [90]

ORP measurements are not temperature compensated because each reaction would have a different correction depending upon the number of electrons transferred The Nernst equation [90, 89] has a coefficient in the log term which depends on both the Absolute Temperature, T, and the number of electrons transferred, n Most industrial applications have a number of different reactions occurring simultaneously Each reaction often has a different number of transferred electrons

A two-electrode system is used to make a measurement (Figure A 19) [92] The metal and reference electrodes are immersed in the solution to be measured, and then the electric potential between the two electrodes is measured Thus, the ORP of the solution is measured A noble metal (a pure, elemental metal) is always used in ORP electrodes because it will not enter into unwanted chemical reactions that can lead to measurement errors The use of a noble metal is important because the ORP value is a function of both the solution's chemicals and the type of metal in contact with that solution (even different noble metals can give different readings in the same solution) The ORP electrode serves as an electron donor or electron acceptor depending upon the test solution A reference electrode is used to supply a constant stable output for comparison Electrical contact is made with the solution using a saturated potassium chloride (KCl) solution [89]

Important factor to consider when making ORP measurements is that they can be pH dependent (as the pH of the solution varies, the potential of the reaction

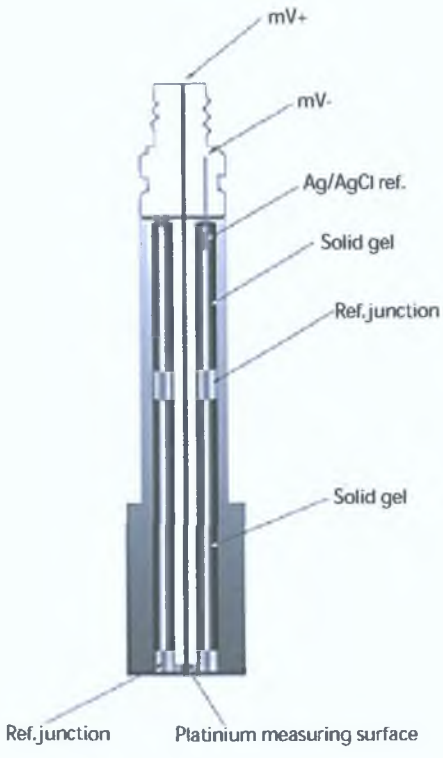

Figure A.19: Schematic of an ORP electrode. [91]

varies).

A.4.2 pH sensor - electrode type

The term pH derives from a combination of p for the word power and H for the symbol of the element Hydrogen. pH is the negative logarithm of the *activity* of hydrogen ions [93].

pH represents the *activity* of hydrogen ions in a solution, at a given temperature. The term activity is used because pH reflects the amount of available hydrogen ions not the concentration of hydrogen ions.

pH sensors are used in medicine for measuring pH of the blood, water treatment, waste water treatment, food industry and other applications that require control and monitoring of the pH.

A sensor usually consists of two electrodes. The sensor measures the electric po-

tential between the pH sensing electrode (which is sensitive just to Hydrogen ions [H+]) and a reference electrode (which provides a constant potential, independent of the concentration of Hydrogen ions [H+]) [93]. Changes in the electric potential between electrodes are caused by differing concentrations of Hydrogen ions [H+].

The construction of the pH electrode is similar to the ORP electrode (Figure A.20) [94].

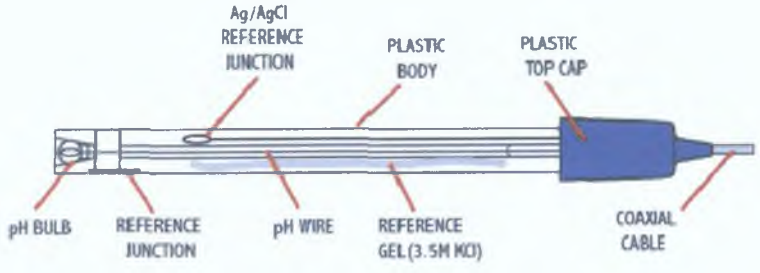

Figure A.20: A pH electrode.

The most common problems experienced with pH electrodes are [94]:

1. Oily and solid coatings requiring frequent removal for cleaning.
2. pH bulb breakage or premature failure as a result of abrasives or solid materials in the solution.
3. Reference junction fouling or plugging requiring frequent removal for cleaning.
4. Reference side contamination as a result of interactions between the silver ions in the electrodes and materials in the solution which interact with silver. This result is short electrode life.

There are three types of bulbs used, spherical bulb, hemi-spherical bulb and flat glass. A flat glass self-cleaning electrode has been designed to solve or minimize the above mentioned difficulties [95]. When the electrode's flat measuring surface is exposed to turbulent flow, the resulting scrubbing action provides a self-cleaning effect in most applications.

For many applications, a single junction reference electrode is used but sometimes unwanted side chemical reactions may lead to erroneous reference signals or to precipitation at the reference junction leading to a short service life [94]. A double junction reference design affords a barrier of protection from such interactions. The safest approach is to use the double junction (see Figure A.20).

A.4.3 pH sensor - Fiber optic chemical sensors

The fiber optic sensor consists of membrane with an appropriate indicator. The indicator changes its optical properties (e.g. absorbance, fluorescence) depending on the analyte. Such an optical signal with information about a sample under test is converted into an electrical signal in an optoelectronic interface. The main part of this interface is the photodetector connected to an electronic circuit [96].

The operating principle of a pH sensor is shown in Figure A.21 [96]. A pulse of light from a light emitting diode (LED) is coupled into an optical fiber and transmitted to a pH sensitive membrane. The membrane changes its absorbance (colour) depending on pH of the sample. If the absorbance is quite low the light is only slightly absorbed which is depicted by almost the same pulse of the light returning to the photodiode. When the pH of the sample is changed the absorbance of the membrane increases so the returning pulse of the light is smaller

Figure A.21: Operating principle of a fiber optic chemical pH sensor.

Light is usually modulated to a square wave so that the measurements are not influenced by an ambient light and in order to increase the signal to noise ratio.

Sensors can be also based on the use of a fluorescence indicator immobilized in the membrane [96].

In this case, the light excites molecules of the indicator which emits light at different wavelength The analyte can, for example, influence the intensity of the fluorescence radiation which is depicted by different amplitude of the returning pulses

Temperature compensation is required

A.4.4 Conductivity sensor

A conductivity sensor is designed to make very rapid, high resolution measurements of the electrical conductivity of water [97] This sensor can be used for investigating turbulent mixing in stratified fluids, measurement of very rapid temperature fluctuations, measurement of mean density profiles, measurement of bubble concentration, measurement of interface position, water quality, plant nutrient status, land reclamation and many other uses

This sensor consists of 4 electrodes (platinum spheres) two drive (current) electrodes and two sense (voltage) electrodes (Figure A 22) [98] The drive electrodes are powered by an alternating voltage, and the alternating current that flows is measured to determine the conductivity The amplitude of the alternating voltage applied to the drive electrodes is controlled by the voltage measured at the sense electrodes Since the sense electrodes are positioned in a low current area of the cell, and this voltage is measured using a high impedance circuit, it represents with high accuracy the strength of the electric field within the cell Using this signal to maintain the cell field strength at a constant, the current that flows at the drive electrodes is proportional to the conductivity of the sample

This additional two sense electrodes are for the compensation of the polarization effects which make a significant error in 2-electrode cells resulting in a lower conductivity reading than the actual value [98]

- Response in seconds
- Temperature compensation is needed
- Non-linear

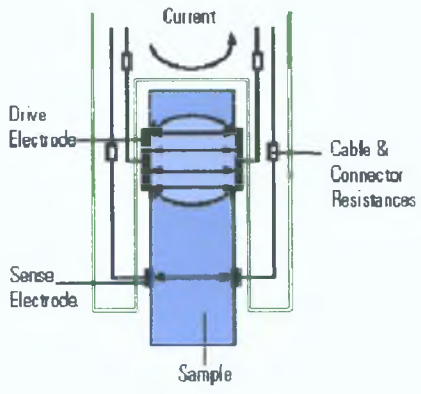

Figure A.22: 4-electrode cell schematic.

A.4.5 Electrochemical dissolved oxygen sensor

They are used primarily for controlling and monitoring aeration processes in wastewater treatment plants, as well as in specific applications such as fish farms, oceanographic research, environmental monitoring, aquaculture, industrial processes [94]. To initiate measurement a polarization voltage is applied between two electrodes. Gold is used as cathode and a silver wire as anode, both placed in an electrolytic system separated from the liquid by a gas permeable diaphragm. Oxygen molecules penetrating the diaphragm are attracted to the positively charged gold cathode. A small electric current is thereby established between the gold cathode and silver anode [99].

Response time less than a minute.

A.5 SAW (Surface Acoustic Waves)

SAW sensors are primarily used to detect gases and VOCs (volatile organic compounds), both halogenated and nonhalogenated, in the vapor phase, semivolatile organic compounds (SVOCs), polychlorinated biphenyls (PCBs), nerve and blister agents, identification of ionic solutions, medical applications (chemical sensors), and industrial and commercial applications (vapor, humidity, temperature, and mass sensors) [100]. Sensors that require no operating power are highly desir-

able for remote monitoring of chemical vapors, moisture, and temperature. Other applications include measuring force, acceleration, shock, angular rate, viscosity, displacement, and flow, in addition to film characterization. The sensors also have an acoustoelectric sensitivity, allowing the detection of ionic contaminants and electric fields. For liquid sensing there is a special class of shear-horizontal surface acoustic wave sensors [100].

Surface acoustic wave devices consist of analogue radio frequency (RF) components operating at frequencies from 30 to about 2500 MHz. They can be used for both wireless identification and sensing. For identification purposes, a SAW transponder (a wireless communications, monitoring, or control device that picks up and automatically responds to an incoming signal derived from the words transmitter and responder) picks up an electromagnetic request signal and stores it until all echoes caused by multipath propagation have died away. Then, a characteristic response is beamed back to the receiver. In radiolink sensors, a physical or chemical quantity influences the propagation properties of the SAW and consequently changes the response pattern of the device [6].

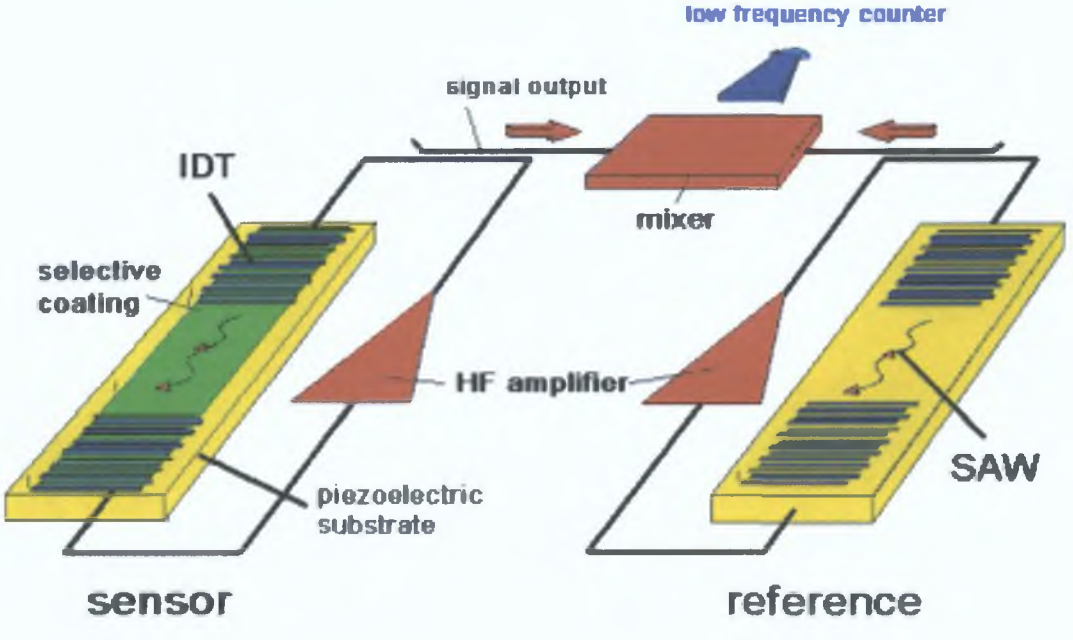

Figure A.23: Measuring Principle of SAW sensor.

The Figure A.23 is an example of a sensing application of SAW in a chemical

sensor where the SAW devices serve as the frequency determining elements of an oscillator circuit [100] One is coated and represents the real sensor, the other is left uncoated and serves as a reference for temperature compensation and generation of a difference frequency by mixing of both oscillator signals The absorption of analytes causes a shift in frequency which is easily counted by a low frequency counter

As the acoustic wave propagates through (Figure A 23) or on the surface of the material, any changes to the characteristics of the propagation path affect the velocity and/or amplitude of the wave Changes in velocity can be monitored by measuring the frequency or phase characteristics of the sensor and can then be correlated to the corresponding physical quantity being measured [100]

Temperature Sensor [100] Surface wave velocities are temperature dependent and are determined by the orientation and type of crystalline material used to fabricate the sensor

An interesting property of quartz (SiO_2) is that it is possible to select the temperature dependence of the material by the cut angle and the wave propagation direction Temperature sensors based on SAW delay line oscillators have millidegree resolution, good linearity, and low hysteresis [100]

Dew Point/Humidity Sensor [100] If a SAW sensor is temperature controlled and exposed to the ambient atmosphere, water will condense on it at the dew point temperature, making it an effective dew point sensor

Appendix B

Results Using Taylor Series

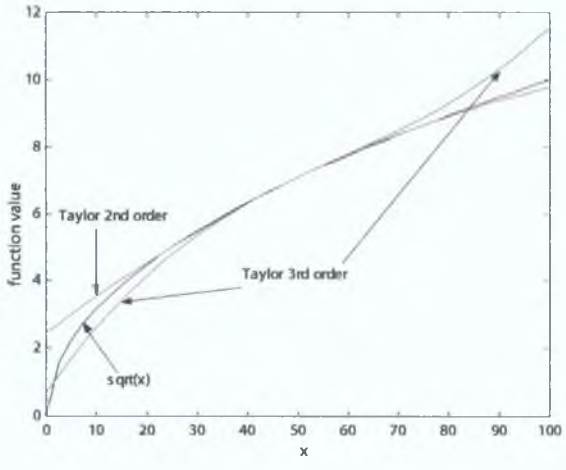

Figure B.1: Estimated function $\sqrt{x}$ in C using Taylor series of the second and the third order, both calculated in 5 points with a reference point at $x = 50$ plotted in MATLAB.

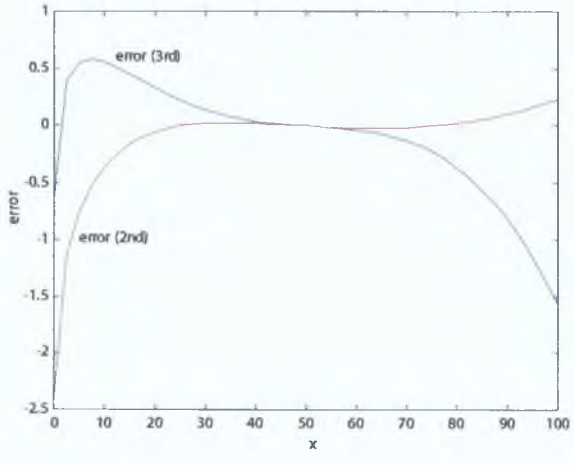

Figure B.2: Approximation error between function $\sqrt{x}$ with a reference point at $x = 50$ plotted in MATLAB.

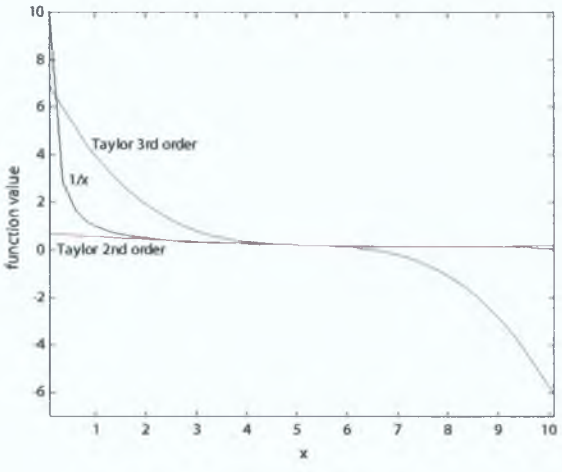

Figure B.3: Estimated function x^{-1} in C using Taylor series of the second and the third order, both calculated in 5 points with a reference point at $x = 5.1$ plotted in MATLAB.

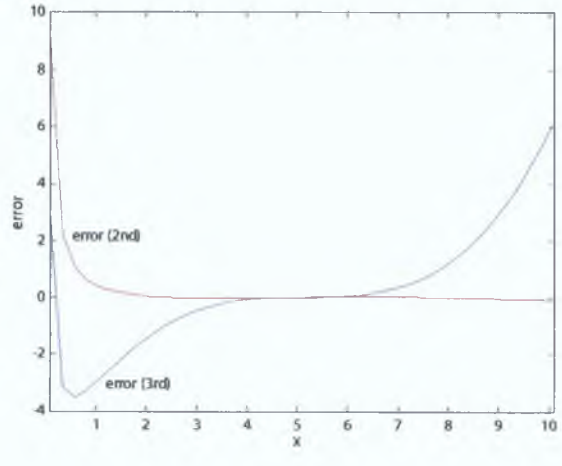

Figure B.4: Approximation error between function x^{-1} and Taylor series with a reference point at $x = 5.1$ plotted in MATLAB.

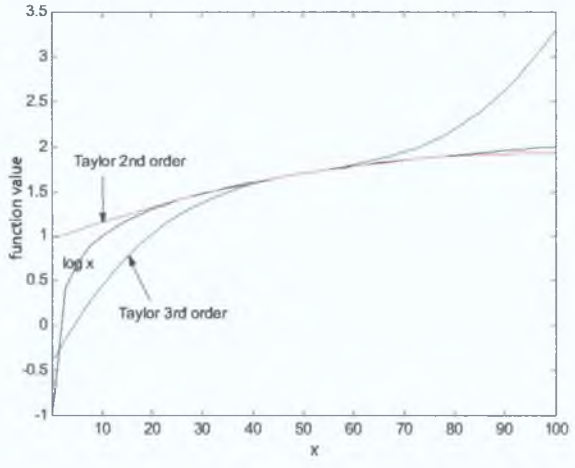

Figure B.5: Estimated function $\log x$ in C using Taylor series of the second and the third order, both calculated in 5 points with a reference point at $x = 51$ plotted in MATLAB.

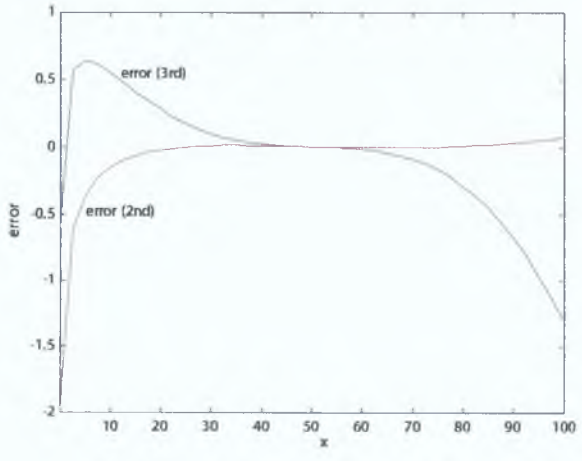

Figure B.6: Approximation error between function $\log x$ and Taylor series with a reference point at $x = 51$ plotted in MATLAB.

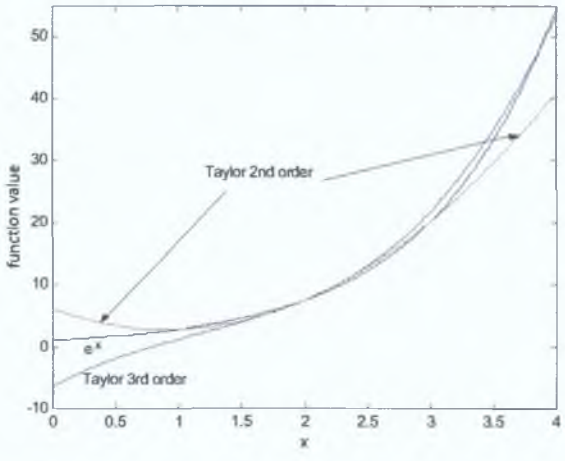

Figure B.7: Estimated function e^x in C using Taylor series of the second and the third order, both calculated in 5 points with a reference point at $x = 2$ plotted in MATLAB.

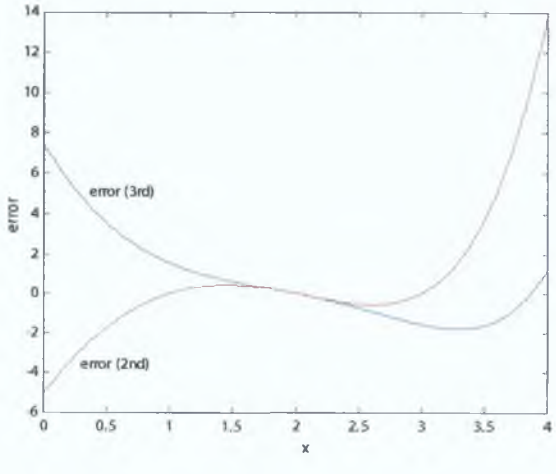

Figure B.8: Approximation error between function e^x and Taylor series with a reference point at $x = 2$ plotted in MATLAB.

Appendix C

Results Using Gaussian Elimination Method

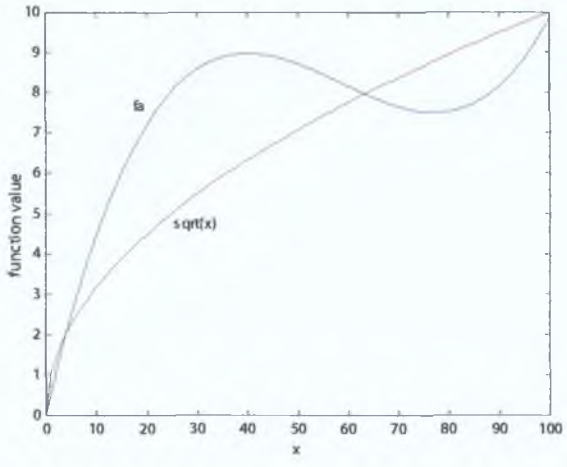

Figure C.1: Estimated function $\sqrt{x}$ using using third order polynomial plotted in MATLAB.

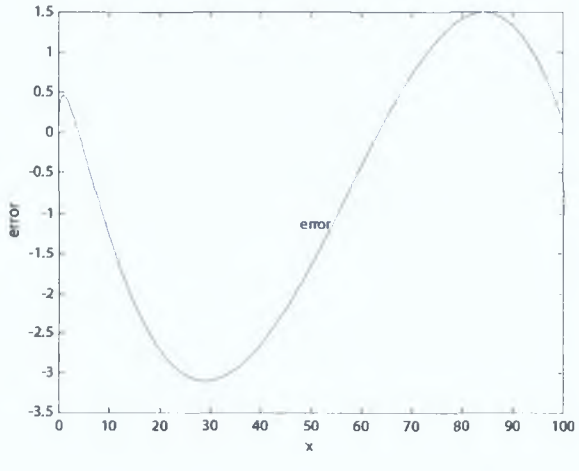

Figure C.2: Approximation error between function $\sqrt{x}$ and third order polynomial plotted in MATLAB.

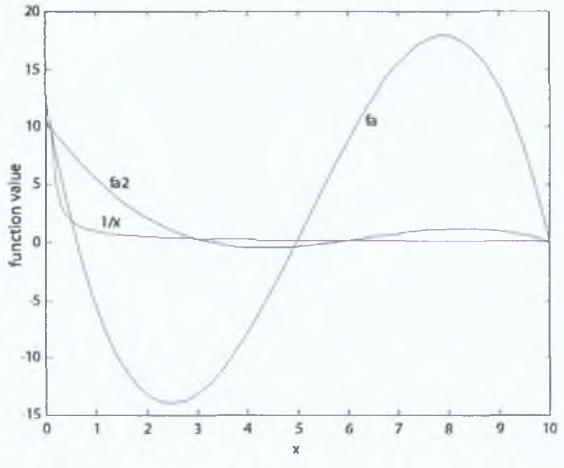

Figure C.3: Estimated function x^{-1} using third order polynomial plotted in MATLAB.

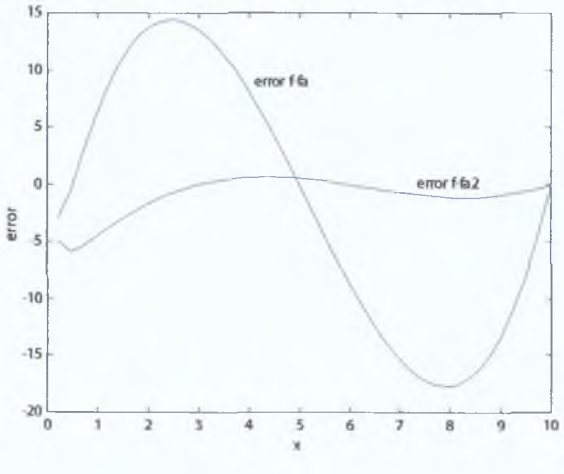

Figure C.4: Approximation error between function x^{-1} and third order polynomial plotted in MATLAB.

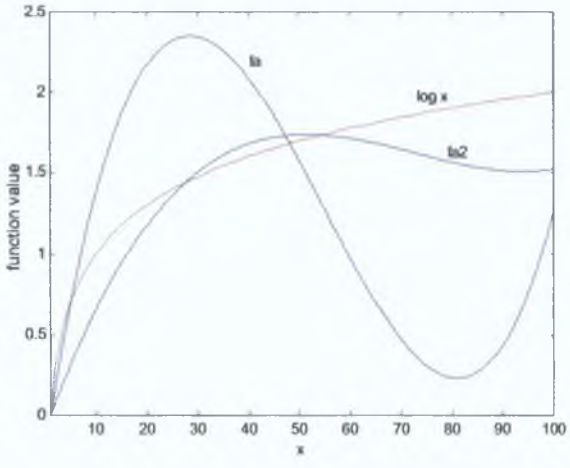

Figure C.5: Estimated function $\log x$ using third order polynomial plotted in MATLAB.

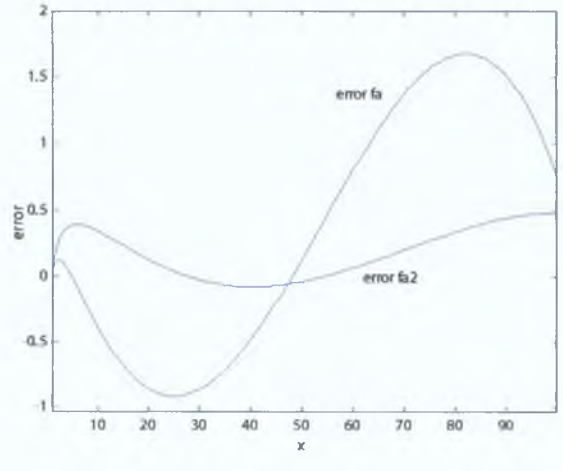

Figure C.6: Approximation error between function $\log x$ and third order polynomial plotted in MATLAB.

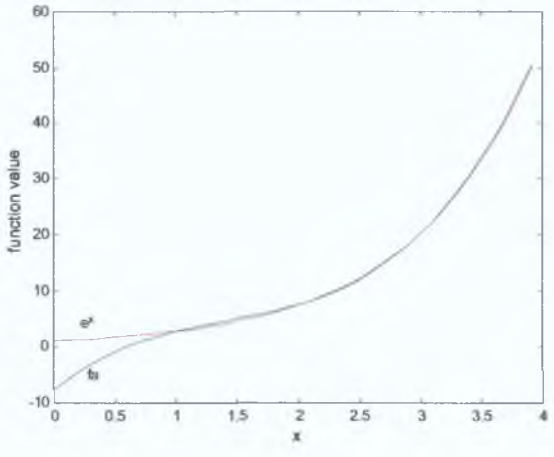

Figure C.7: Estimated function e^x using third order polynomial plotted in MATLAB.

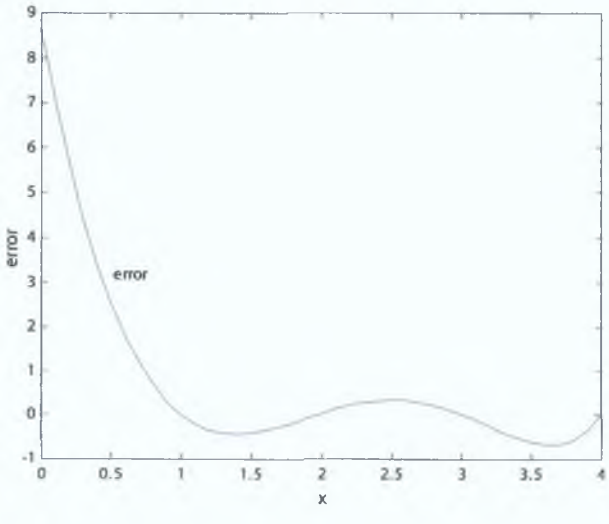

Figure C.8: Approximation error between function e^x and third order polynomial plotted in MATLAB.

Appendix D

Results Using Lagrange Interpolation

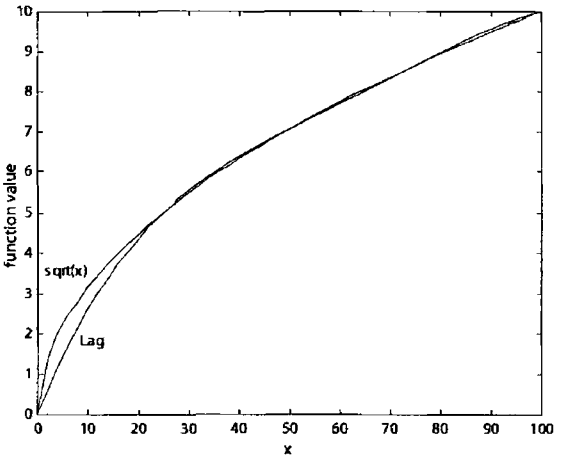

Figure D 1 Estimated function $\sqrt{x}$ using Lagrange interpolation plotted in MATLAB

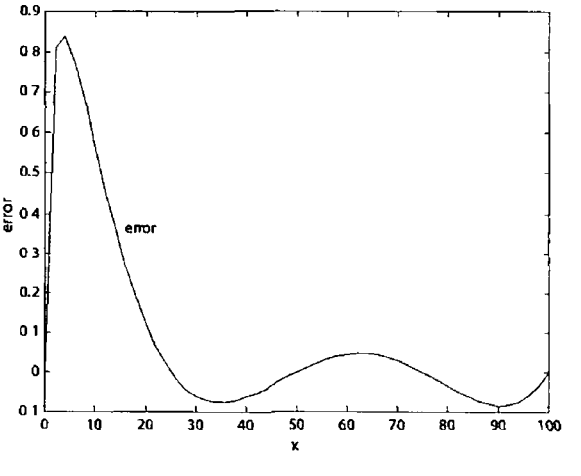

Figure D 2 Approximation error between function $\sqrt{x}$ and and Lagrange interpolation plotted in MATLAB

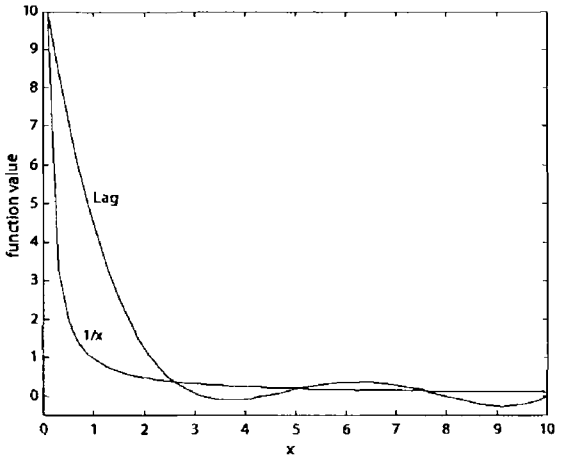

Figure D 3 Estimated function x^{-1} using Lagrange interpolation plotted in MATLAB

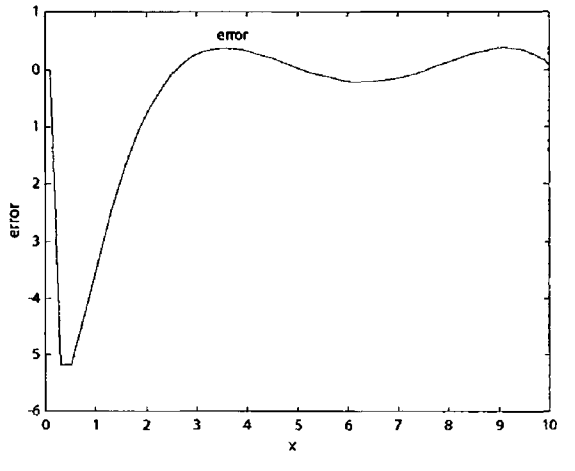

Figure D 4 Approximation error between function x^{-1} and Lagrange interpolation plotted in MATLAB

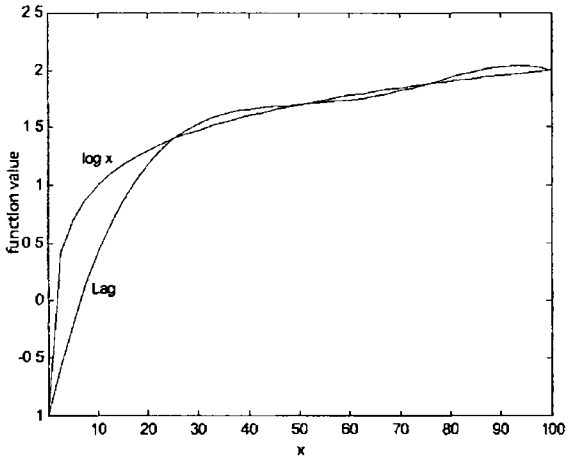

Figure D 5 Estimated function $\log x$ using Lagrange interpolation plotted in MATLAB

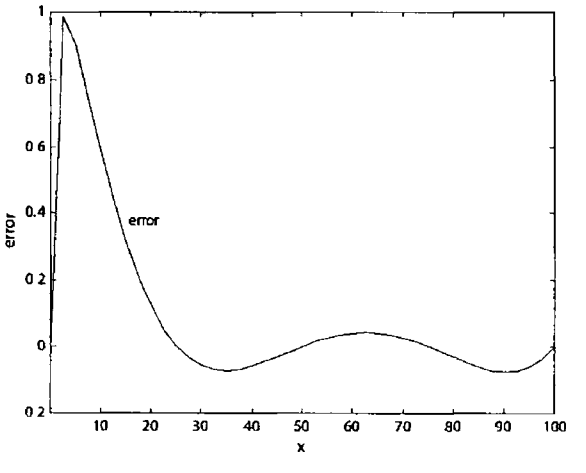

Figure D 6 Approximation error between function $\log x$ and Lagrange interpolation plotted in MATLAB

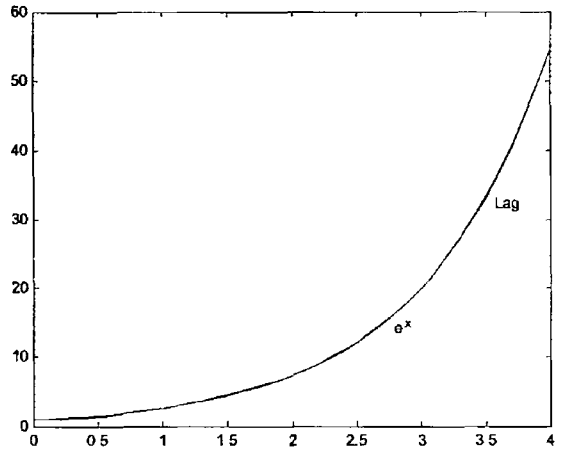

Figure D 7 Estimated function e^x using Lagrange interpolation plotted in MATLAB

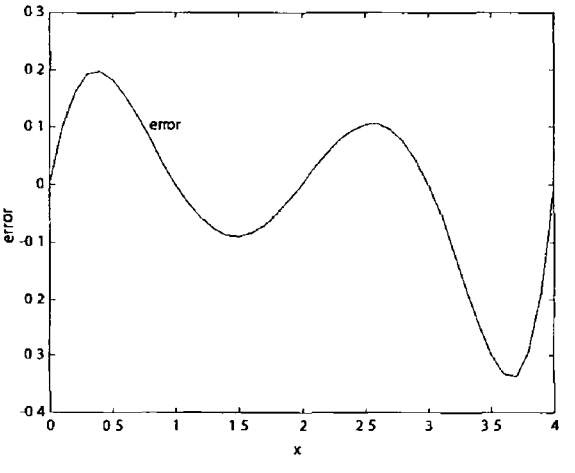

Figure D 8 Approximation error between function e^x and Lagrange interpolation plotted in MATLAB

Appendix E

Results PPC

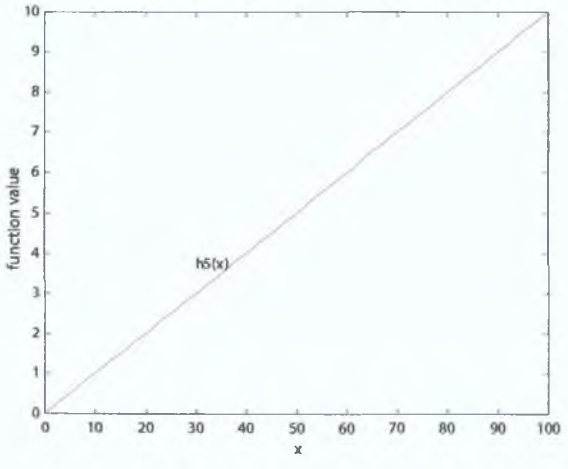

Figure E.1: Corrected sensor TF $h_5(x)$ of the sensor TF $\sqrt{x}$ using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 100, 50, 75, 25\}$).

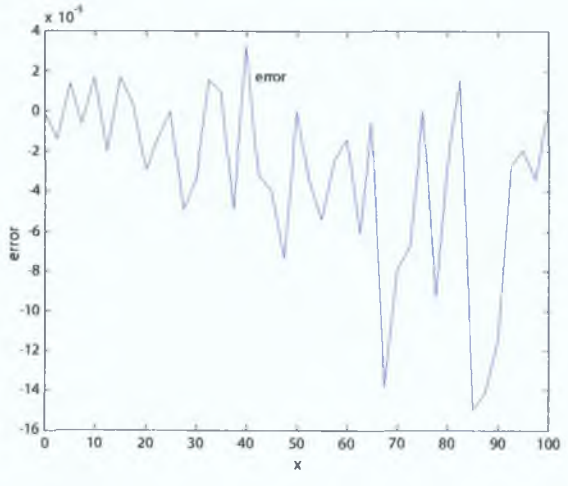

Figure E.2: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0, 100, 50, 75, 25\}$).

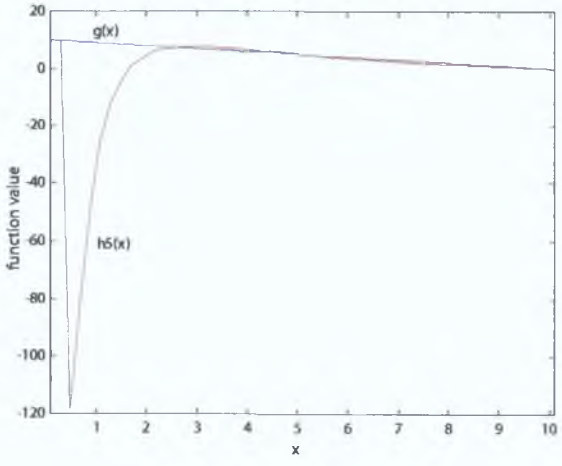

Figure E.3: Corrected sensor TF $h_5(x)$ of the sensor TF x^{-1} using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0.1, 10.1, 2.6, 5.1, 0.3\}$).

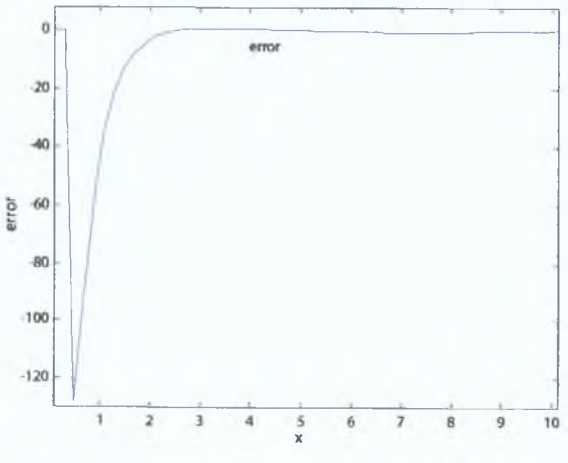

Figure E.4: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0.1, 10.1, 2.6, 5.1, 0.3\}$).

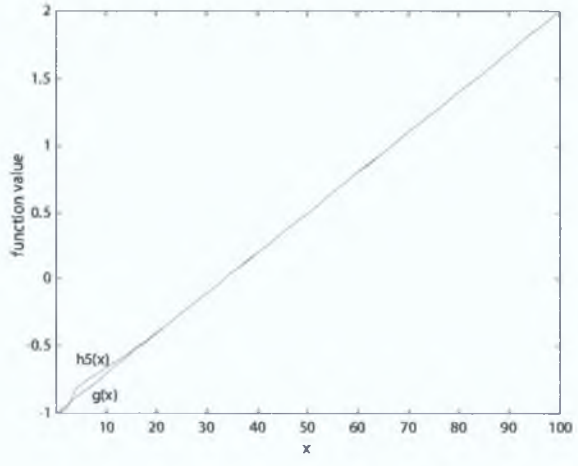

Figure E.5: Corrected sensor TF $h_5(x)$ of the sensor TF $\log x$ using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0.1, 100.1, 50.1, 75.1, 25.1\}$).

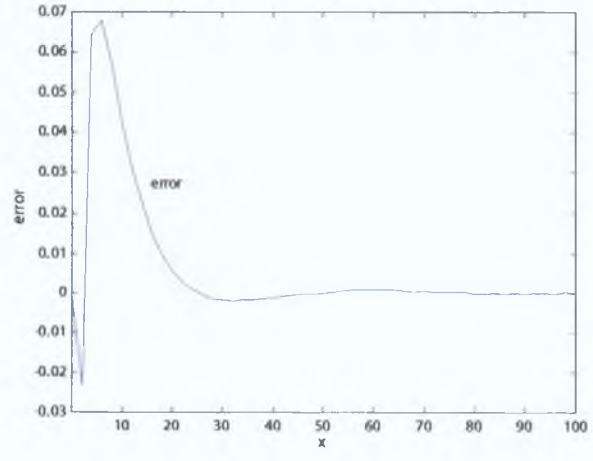

Figure E.6: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0.1, 100.1, 50.1, 75.1, 25.1\}$).

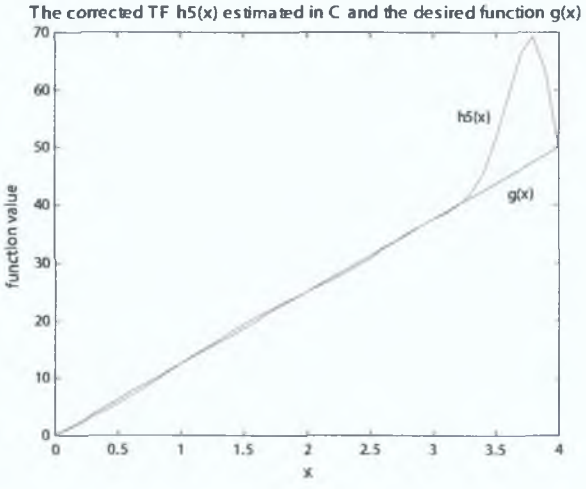

Figure E.7: Corrected sensor TF $h_5(x)$ of the sensor TF e^x using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 4, 2, 3, 1\}$).

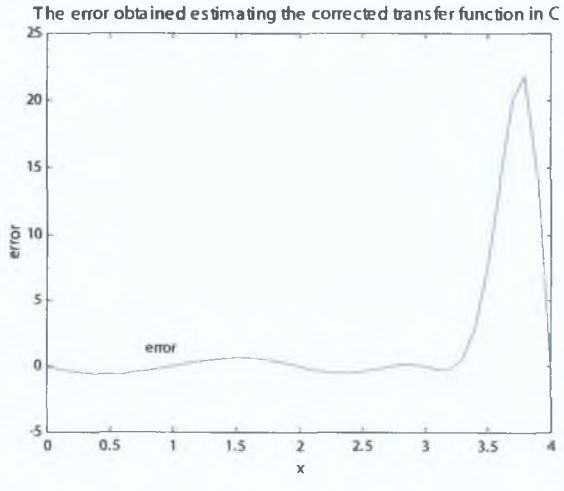

Figure E.8: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0, 4, 2, 3, 1\}$).

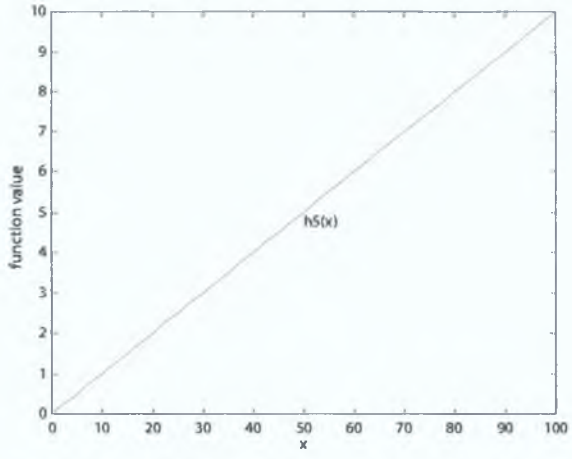

Figure E.9: Corrected sensor TF $h_5(x)$ of the sensor TF $\sqrt{x}$ using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 100, 50, 25, 75\}$).

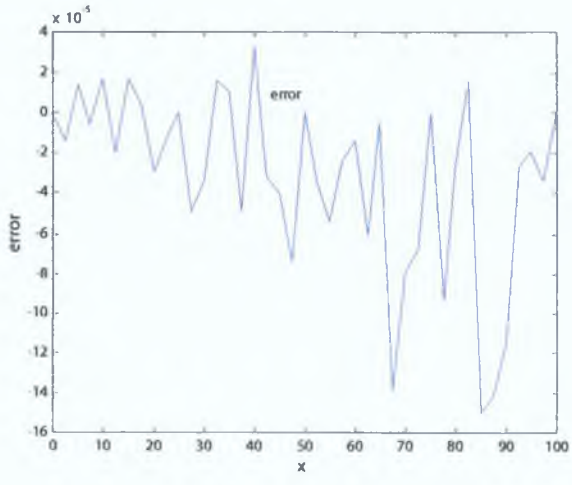

Figure E.10: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0, 100, 50, 25, 75\}$).

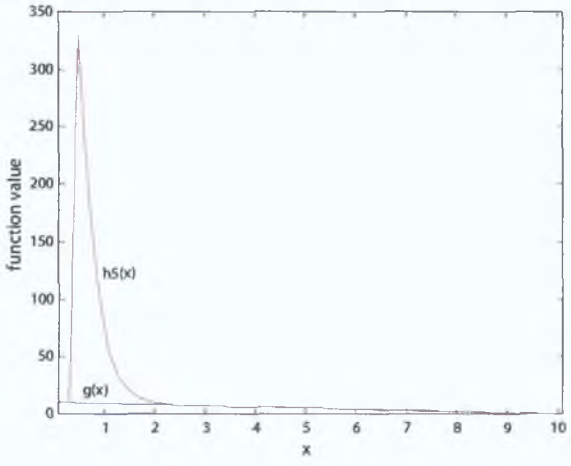

Figure E.11: Corrected sensor TF $h_5(x)$ of the sensor TF x^{-1} using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0.1, 10.1, 2.6, 0.3, 5.1\}$).

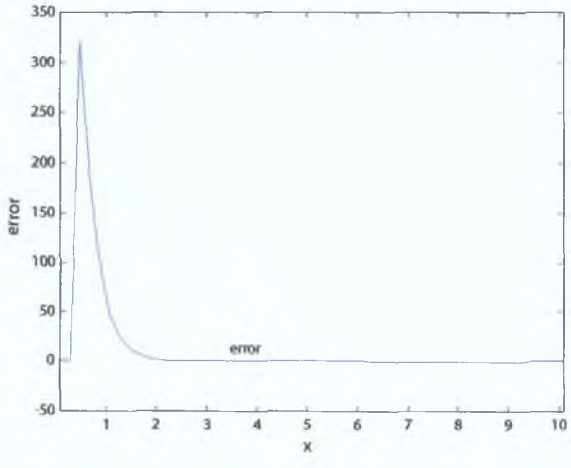

Figure E.12: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0.1, 10.1, 2.6, 0.3, 5.1\}$).

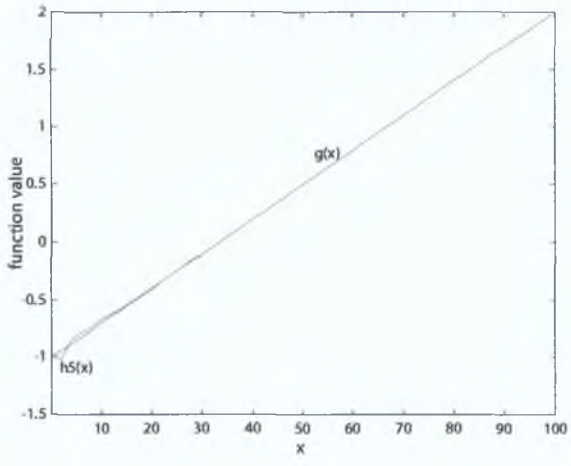

Figure E.13: Corrected sensor TF $h_5(x)$ of the sensor TF $\log x$ using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0.1, 100.1, 50.1, 25.1, 75.1\}$).

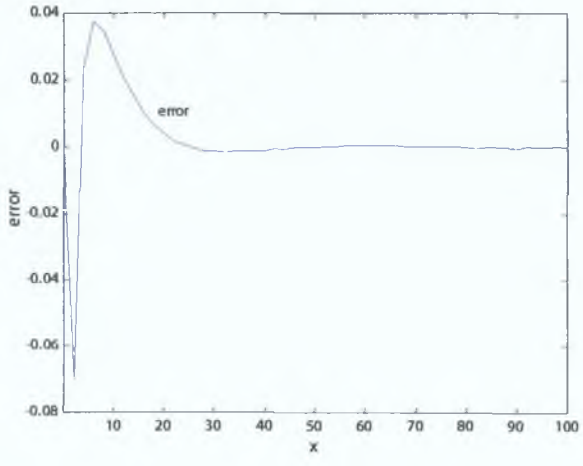

Figure E.14: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0.1, 100.1, 50.1, 25.1, 75.1\}$).

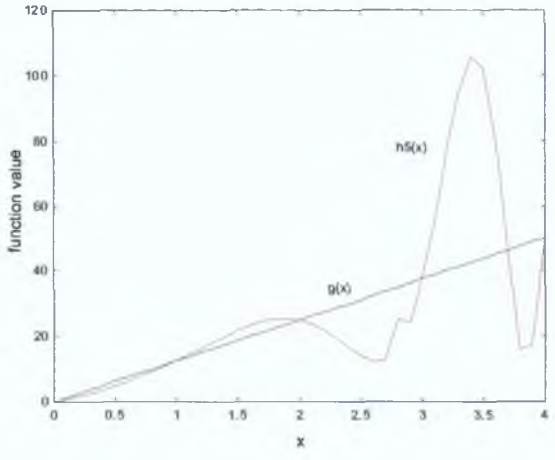

Figure E.15: Corrected sensor TF $h_5(x)$ of the sensor TF e^x using PPC method in 5 points and desired sensor TF $g(x)$ plotted in MATLAB ($x = \{0, 4, 2, 1, 3\}$).

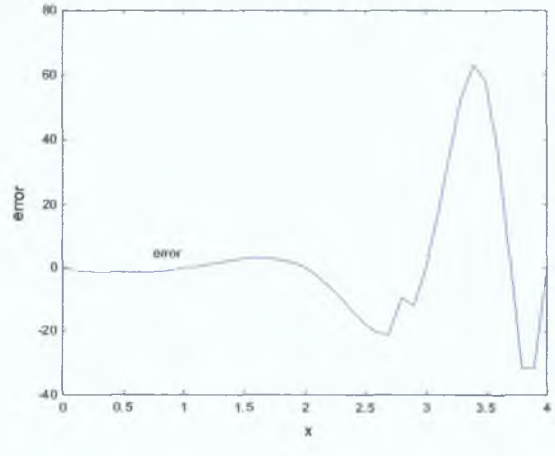

Figure E.16: The error $\varepsilon = h_5(x) - g(x)$ for 5 calibration measurements plotted in MATLAB ($x = \{0, 4, 2, 1, 3\}$).