Ameliorating integrated sensor drift and imperfections: an adaptive "neural" approach

Tong Boon Tang

Abstract

The work carried out towards this thesis examines the suggestion that local pre-processing and early classification of high-dimensional sensory signals can be achieved effectively by an artificial neural network (ANN). The multisensor microsystem for a project named "Integrated Diagnostics for Environmental & Analytical Systems (IDEAS)" is used as an example for this study. Four types of electrochemical sensors are implemented and calibrated. In our testbench experiments, the sensory signals are found to experience some stochastic randomness and drift during operation. Therefore, the ANN must be adaptive to allow auto-calibration of the sensors.

Existing ANN algorithms to fuse sensory signals and hardware-amenable neural algorithms are reviewed. The Continuous Restricted Boltzmann Machine (CRBM) is chosen as a suitable candidate. The CRBM is further developed in this thesis to facilitate on-line learning without experiencing *Catastrophic Interference (CI)* - a known problem in associative memory based models.

The CRBM is examined in two separate simulations. The first simulation evaluates the modelling capability of the CRBM while the second simulation focusses on the adaptation of the CRBM to sensor drift in a dynamic environment. The results suggest that the CRBM is able to model high-dimensional, non-Gaussian data distributions with overlapping areas. The CRBM is also compared favourably, in terms of robustness against sensor drift, with trained but subsequently non-adaptive neural models. The thesis also investigates the optimal architecture size and learning rate for the CRBM. Finally, the limitations of the CRBM are studied. The learning rate is identified as the key factor in determining the feasibility of CRBM tracking sensor drift in a dynamic environment.

Declaration of originality

I hereby declare that the research recorded in this thesis and the thesis itself was composed and originated entirely by myself in the Department of Electronics and Electrical Engineering at The University of Edinburgh.

Tong Boon Tang

Acknowledgements

Contents

List of figures

List of tables

Acronyms and abbreviations

ADC	Analogue-to-digital converter
AI	Artificial intelligence
ANN	Artificial neural network
ART	Adaptive Resonance Theory
BER	Bit error rate
CE	Counter electrode
CI	Catastrophic Interference
CNN	Cellular neural network
COPDS	Chronic pulmonary diseases
CPU	Central processing unit
CRBM	Continuous Restricted Boltzmann machine
DAC	Digital-to-analogue converter
DFA	Discriminant Factoral Analysis
DO	Dissolved oxygen
DS-SS	Direct-sequence spread spectrum
DSP	Digital signal processor
ECG	Electrocardiogram
EDA	Electronic Design Automation
EEG	Electroencephalogram
GA	Genetic Algorithm
GI	Gastro-intestinal
GUI	Graphical user interface
IC	Integrated circuit
IDEAS	Integrated Diagnostics for Environmental & Analytical Systems
IFFN	Identification-friend-foe-neutral
ISF	Intelligent sensor fusion
ISFET	Ion-sensitive field effect transistor
JDL	Joint Directors of Laboratories
KNN	k-nearest neighbour

LDA	Linear discriminant analysis
LIAP	Lab-in-a-Pill
LSB	Least significant bit
LVQ	Learning Vector Quantization
MCD	Minimising Contrastive Divergence
MCU	Microcontroller unit
MLP	Multi-Layer Perceptrons
MSE	Mean-square error
PC	Personal computer
PCA	Principal component analysis
QCR	Quartz crystal resonator
RBF	Radial Basis Functions
RE	Reference electrode
REFET	Ion-insensitive field effect transistor
RF	Radio frequency
RNN	Recurrent neural network
SFI	Supervised fuzzy isodata
SLP	Single layer perceptron
SNR	Signal-to-noise ratio
SoC	System-on-Chip
SOM	Self-organizing map
TDS	Total level of dissolved solids
VLSI	Very Large Scale Integration
WE	Working electrode
WIMS	Wireless Integrated MicroSystems
WINS	Wireless Integrated Network Sensors

Chapter 1
Introduction

1.1 Background and Motivations

This project brings together sensor integration and neural computation. As the focus is on developing a system to deal with early sensor pre-processing, this section will provide a brief overview of the sensor technology and the needs to pre-processing.

1.1.1 Integrated multisensor microsystems

Since the late 1970s, silicon has been identified as a promising sensor substrate because it permits the integration of the sensing element and the signal processing circuit on a single chip, and allows batch-fabrication process. The types of transduction available in silicon include radiant, mechanical, thermal, magnetic and chemical signal conversion from a wide range of measurands (as summarised in Table 1.1). While silicon displays sufficient physical effects to most types of measurands, some modification to the silicon surface is required to enhance the sensor sensitivity to detect other measurands. For example, to make a magnetic field sensor, a thin Nickel-iron layer must be deposited on top of the silicon substrate; to create a pH sensor, the metal gate of a silicon field effect transistor is replaced by an ion-sensitive membrane that has affinity for protons.

Typically, the transduction has limited sensitivity and sensor outputs are in analogue format, making them particularly susceptive to noise when transmitted through long wires. The signal-to-noise ratio (SNR) can be improved considerably by integrating the sensor and its signal conditioning circuit onto the same substrate. Such a sensor is better known as an "integrated sensor". It achieves better SNR by using local amplification of the raw sensor signal. Furthermore, an integrated sensor requires shorter interconnects and hence suffers less parasitic capacitance effects. From the manufacturing point of view, it is cheaper to package than those that use a separate signal conditioning chip.

1

Signal domains	Example measurands	Physical effects
Radiant	light intensity, wavelength, polarization, phase	photoconductivity, photovoltaic, photo-electric and photomagneto-electric effects
Mechanical	force, pressure, vacuum, flow, tilt, thickness	piezoresistivity, lateral photoelectric and lateral photovoltaic effects
Thermal	temperature, temperature gradient, heat, entropy	Seebeck effect, temperature dependence of conductivity and junction, Nernst effect
Magnetic	field intensity, flux density, permeability	magnetoresistance, Hall and Suhl effects
Chemical	concentration, toxicity, pH, reduction potential	ion-sensitive field effect

Table 1.1: *List of signal domains with their examples and physical effects in silicon [2].*

Moreover, on-chip signal processing available in integrated sensors can also offer the following advantages [2]:

- Improved sensor characteristics

  1. Non-linearity; most sensors demonstrate some non-linearity. Using an on-chip feedback system or a look-up table can improve linearity.

  2. Cross-sensitivity; most sensors have an undesirable sensitivity to other measurands (e.g. temperature and interfering ions). This can be counteracted by incorporating relevant sensors and circuits into the same chip.

  3. Offset; some sensors do not give a zero output when the readout should be zero. The offset can be compensated by using either a bias circuit with some laser-trimming resistors or a digital-to-analogue converter (DAC) at the input stage of signal amplification.

  4. Parameter drift; sensor parameters tend to be a function of time. This can lead to a change in sensitivity, offset, linearity, etc. If an accurate current or voltage source is incorporated on-chip, then well defined values of measurands can be generated locally for the purpose of auto-calibration.

  5. Frequency response; proper feedback methods can increase the sensor operating bandwidth.

- Signal conditioning and formatting

  1. Analogue-to-digital conversion; nearly all sensors generate analogue signals.

Digital signals have very good noise immunity. Therefore, on-chip analogue-to-digital conversion is very desirable.

2. Impedance transformation; most sensors have high internal impedance. While measuring their output voltage *directly*, some current will inevitably be drawn and the actual voltage amplitude will thus be attenuated. Using a buffer at the sensor output stage can prevent such problem.

3. Output formatting; different sensor principles result in different output formats (conductivity, capacitance, voltage etc.). With on-chip signal processing, it is possible to convert all measurements into one single format, facilitating multisensor integration.

4. Output averaging; a high bandwidth databus is required when many redundant sensors are used. On-chip electronic circuits can employ a median filter to minimise the required bandwidth.

- High-level signal processing
  Integrating more sophisticated signal processing functions on the transducer chip is in parallel to the trend of Very Large Scale Integration (VLSI) technology [3]. Functions such as binary classification, pattern recognition and sensor fault detection can be implemented onto a microprocessor and/or a digital signal processor (DSP). A power saving scheme can also be implemented to improve electrical energy efficiency.

In present days, multiple sensors are integrated in the same chip to produce *integrated multisensor microsystems*. Using a huge number of identical/disparate sensors, as inspired by biological systems, can improve the microsystem robustness and provide information on various relevant attributes of a target entity. In addition to multiple integrated sensors, the microsystems also incorporate some micropower integrated circuits and, in some cases, wireless communications for a wide range of applications [4]. For instance, the *Wireless Integrated Network Sensors* (WINS) [5] are deployed for surveillance applications in place of more expensive fully-wired systems. Limited by a narrow bandwidth, sensor signals are processed locally. Events of interest provoke an alarm to a nearby basestation for further decision making. Local data fusion (typically classification) uses an artificial neural network (ANN) because it allows data fusion at all levels (signal, pixel, feature and symbol) [6].

Another two similar works on ubiquitous low-cost, low-energy sensor arrays are the *SMARTDUST* and the *picoNode*. Each Dust [7] is built from off-the-shelve components (namely, the sensors, the microprocessor and the battery), and packaged into one cubic centimeter and powered by an on-chip solar cell or combustion fuel. Dusts communicate to each other via laser or radio frequency (at 916MHz) signals, and the range can be over 20m. Likewise, each picoNode [8] has a dimension of one cubic centimeter and a weight less than 100g. To be deployable in many hundreds, each picoNode uses ultra-low power (less than 100 microwatts) to eliminate the need of frequent battery replacement. The communication protocol stack is particularly optimized to minimise power dissipation by making a trade-off between the amount of communication and computation, as well as minimising the required overhead.

As described, integrated multisensor microsystems today, are more complicated than ever, and naturally, with increasing number of available measurements, there is a commensurate demand for a new *intelligent sensor fusion* (ISF) technique. Ideally, this technique should be hardware-amenable and compute at a power level in microwatts. Additionally, it should be robust against stochastic noise and sensor drift. This thesis sets out to search for such technique in the field of ANNs, and apply it in a real-world application - the *IDEAS*. Section 1.1.2 introduces the *IDEAS* project and defines the specifications for the required data fusion.

1.1.2 *IDEAS*: the Lab-in-a-Pill

The Integrated Diagnostics for Environmental & Analytical Systems (*IDEAS*) project is collaborative research undertaken by the Universities of Edinburgh, Glasgow and Strathclyde, and the Institute for System Level Integration, Livingston. It aims to bring together the Lab-on-a-Chip[1] and System-on-Chip[2] (SoC) technologies to build powerful measurement and monitoring tools for use in environmental and biomedical applications. The main target application is an ingestible pill (named "Lab-in-a-Pill" (LIAP)) capable of traversing the digestive system and sending sensor data to a portable base

---

[1]Lab-on-a-Chip is a technology that intends to integrate complex laboratory sensors and sample-handling capabilities onto a glass or silicon plate [9].

[2]SoC is an electronics design methodology that integrates data analysis, instrumentation and communication capabilities onto a single piece of silicon [10]. The re-use of pre-developed electronic circuit modules is an essential component of the SoC concept.
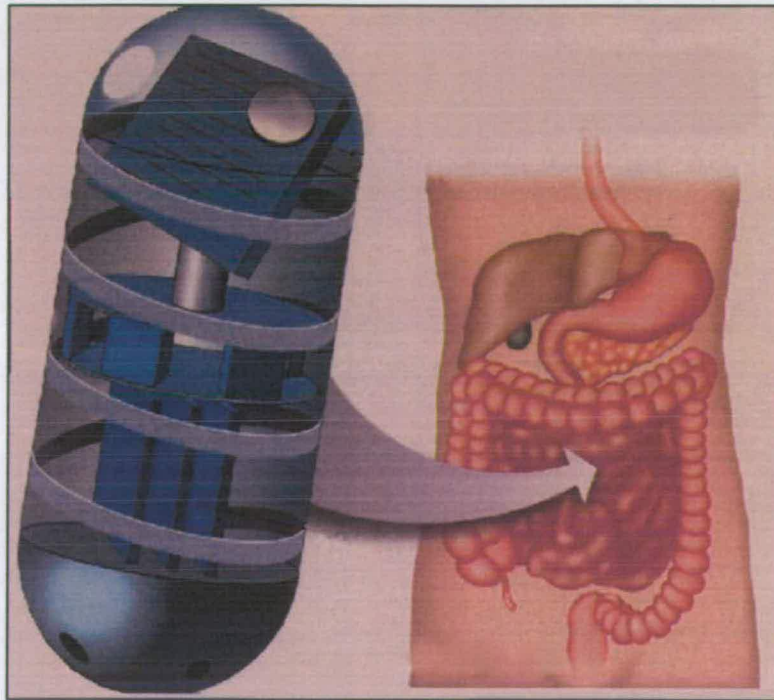
Figure 1.1: *An artistic illustration of a Lab-in-a-Pill traversing the gastrointestinal tract [Photo courtesy of Glasgow University].*

station worn by the patient. Other potential applications range from environmental monitoring and industrial inspections to sensor buoys and localised drug delivery in the body. A similar work has been performed by Given Imaging Ltd. in Israel and their product is now available in the consumer market, named 'PillCam^{TM}Capsule Endoscopy' [11]. Each capsule has miniature colour video cameras on both ends and transmits images of the gastro-intestinal (GI) tract to a data recorder worn on a belt around the waist of a patient. Once the capsule is removed from the patient, the data recorder downloads the video images to a designed workstation, from which the physician views and assesses the results for diagnosis purpose.

The *IDEAS* project involves the design of a capsule and a portable base station system. The capsule measures the conditions in a digestive system with an array of integrated microelectronics sensors. Compared to ones that are glass based, these silicon sensors can be smaller in size and are cheaper to fabricate. However, some (e.g. pH and dissolved oxygen sensors) tend to be less stable and reliable because of corrosion on the encapsulation layer of their miniature on-chip reference electrode(s). When extended

over a length of time, the encapsulation layer will break eventually and result in the sensors malfunctioning.

Moreover, there is a significant variation in some of the sensor intrinsic parameter(s), due to the current fabrication process. For instance, the threshold voltage of pH sensors from one single batch fabrication can vary from 0.28V to 1.88V. Such magnitude of variation means that it is difficult to use a generic signal conditioning circuit for all (functionally-identical) sensors without compromising the sensor sensitivity, given a limited voltage supply. To achieve maximum sensitivity, sensors need to be calibrated individually but this is an expensive practice in terms of time and labour.

Another problem in this particular application is *biofouling*. Biofouling refers to the gradual accumulation of waterborne organisms (such as bacteria and protozoa) on the surfaces of engineering structures in water, which contribute to a decrease in the sensory sensitivity. A compelling solution is to incorporate more redundant sensors on chip to increase the robustness of the measurements. To minimise the required databus bandwidth, binary classification with an ANN is thus adopted as a local data fusion approach.

Based on the aforementioned technical challenges, the specifications for the ANN can be summarised as below:

1. perform binary classification as a form of data fusion

2. ameliorate the imperfection due to instrinsic device variation

3. be robust against stochastic noise

4. autocalibrate against sensor drift

5. operate with analogue signals directly

6. consume low electric energy

7. require simple computation only and be hardware-amenable

8. have a small silicon footprint.

1.2 Objectives of this study

At core, this project sets out to explore the suggestion that

> *local pre-processing and early classification of real, high-dimensional, drifting sensory signals can be achieved effectively by a hardware-amenable neural system.*

This will be examined in the context of the *IDEAS* project. It should be equally applicable to other miniature, power-aware multisensor microsystems.

This project sub-objectives are follows:

1. To perform a literature review on available hardware-amenable ANNs and identify suitable neural system(s) for the aforementioned specifications. Necessary modification to the ANN architecture and training methodology will then be proposed.

2. To obtain sensor models based on actual calibration together with some buffer solutions. This is necessary because real-world measurements are often rare and expensive. Having sensor models allows an unlimited amount of available data for training and testing the neural system.

3. To examine the modelling ability of the proposed neural system with (a) simple high-dimensional but overlapping clusters (generated from obtained sensor models), and (b) non-Gaussian, interleaving two-dimensional clusters (generated artificially). Optimization for training will also be examined in terms of system size (number of neurons) and training time.

4. To examine the adaptivity of the proposed ANN in a dynamic environment where the integrated sensors experience stochastic drift. The drifting data will be obtained through test bench experiment. Limitations of the proposed neural system will be identified.

1.3 Thesis plan

This thesis can be generally divided into two parts. The first part, comprising Chapters 3 and 4, presents the preparatory work for subsequent simulations. This includes the derivation of the sensor models and the development of the neural system. Chapters 5 and 6, as the second part of the thesis, present the simulations with both artificial and real datasets. Following the introduction, the contents of the subsequent chapters can be summarised as follows:

- **Chapter 2** introduces intelligent sensor fusion and reviews (i) ANNs applied in chemical sensor fusion, (ii) current counter-drift approaches and (iii) available hardware-amenable neural systems.

- **Chapter 3** first explains the basic operation principles of the integrated sensors, then presents the interface circuit designs for the integrated sensors and derives the sensor models.

- **Chapter 4** addresses the development of the neural system and explains its training methodology. Practical issues are also discussed.

- **Chapter 5** presents the binary classification simulations on the sensor models and on artificial datasets. The former involves a high-dimensional and overlapping clusters problem, while the latter looks at a non-Gaussian, interleaving two-dimensional clusters problem.

- **Chapter 6** presents the binary classification simulations on a real drifting dataset. As in the previous chapter, two types of cluster problems are examined. Results are compared with two benchmarks (both linear and non-linear neural classifiers).

- **Chapter 7** first summarises the work done, then draws some conclusions based on the simulation results and eventually proposes several possible progressions of this research.

Chapter 2
Literature review

This chapter aims to provide the necessary background to this work and a review of relevant literature in order to identify a possible solution to the application (*IDEAS*) problem. Firstly, Section 2.1 introduces the basic concepts and the four different mechanisms of intelligent sensor fusion. This thesis proposes to use a "neural" approach as a form of data fusion. Therefore, research on applying artifical neural network (ANN) in electrochemical sensor fusion is reviewed in Section 2.2. Section 2.3 presents current solutions available to the drift problem. The chapter is concluded with a survey on available neural hardware in analogue VLSI (aVLSI) technology in Section 2.4.

2.1   Intelligent sensor fusion

2.1.1   An introduction

Integrated sensor processing in this work falls into the study of intelligent sensor fusion (ISF). To date, extensive work in ISF has been done primarily in relation to robotics and military applications. Example applications include the Honda humanoid robot [12], the interactive museum tour-guide robot [13], an automated air defense system [14] and an identification-friend-foe-neutral (IFFN) system [15]. Whilst the actual implementation may be different, the basic concepts of ISF are equally applicable to emerging biomedical applications such as the *IDEAS* which incorporates different electrochemical sensors in a pill.

As defined in [16], *sensor fusion* is a process which autonomously gathers observations (e.g. temperature variation, dissolved oxygen concentration and images of a moving object) from multiple sensors and combines them into a single, coherent percept. *Intelligent sensor fusion* allows the sensor fusion mechanism to **adapt** itself to major environment changes and sensor malfunctions, and can determine its own sensing strategies for observing the percept in order to maintain an efficient use of shared sensing resources.

There are three common motivations to introduce any form of ISF to a sensing system:

1. *to reduce uncertainty in decision making*

   Uncertainty arises in several situations; for instance, when some measurements are missing (e.g. occlusions), when the sensor cannot measure all relevant attributes of the entity (e.g. a pH sensor cannot measure thermal radiation), and when the observation is ambiguous (e.g. a change in pH reading may be due to thermal variation or actual change in H^+ ion concentration). Such uncertainty can be reduced by introducing complementary sensors (e.g. a thermocouple in this case). Note that however ISF cannot improve any imprecision due to the sensor itself. Imprecision is limited by the sensor intrinsic sensivity [17].

2. *to improve robustness in measurement*

   In a real world, each measurement is subjected to stochastic noise, sensor drift and failure. The noise can be caused by many different sources, for instance, the interfering ions due to limited selectivity in an electrochemical sensor, or the high frequency coupling noise from a radio frequency (RF) circuit in a monolithic device. The sensor drift, which is defined as a slow rate of change in a signal output value not correlated with changes in the physical parameter [18], is caused by instability in the sensing reference point. These problems are exacerbated in a harsh environment where the sensor failure rate is high. By implementing redundant sensors, the chance of obtaining reliable measurements can be improved. Furthermore, some form of intelligence can be incorporated to provide the ability to adapt to sensor drift and/or allow the performance to degrade *gracefully* before a complete sensor failure.

3. *to incorporate contextual knowledge*

   Several multisensor integration architectures have been proposed in the literature. Among these, there are the Joint Directors of Laboratories (JDL) model [6] and the NAVLAB system [19]. The JDL model includes sensor selection and controller modules to tailor sensing configuration and fusion process in response to the current environment. Similarly, the NAVLAB system employes a centralised database called the "CODGER" system to structure the information flow from different distributed sensing modules. In essence, these modules/systems are employed to extract the most accurate information in the presence of the common

sensor problems - random noise, drift and failure.

Driven by recent advancements in microsystem technology, applications of ISF are becoming more widespread and interests grow beyond military and robotic based systems. At the Center for Wireless Integrated MicroSystems (WIMS, University of Michigan, USA), biomedical sensors have been developed for cochlear prosthesis. An array of implantable microelectrodes were used to sense biopotentials and neurochemicals in the local nervous system [20]. The simultaneous neural recording was coordinated by a controller and the digitised measurement was transferred, via a wireless RF link, to an external receiver for further data fusion [21].

In contrast to the aforementioned system which employed only a *single* type of sensor, van Putten *et al.* [22] illustrated an example of local data fusion from four different types of sensors to obtain a complete set of attributes of the event of interest - asthma and chronic pulmonary diseases (COPDS). Peak expiratory flow, temperature, pressure and relative humidity were measured with integrated silicon Micro-Electro-Mechanical Systems (MEMS). All measurements were fused through a central module where data processing and data storage were performed, before passing them to a display and a transmitter.

Besides biomedical applications, ISF has also been utilised for indoor environment monitoring, remote sensing and fault detection in machinery. In [23], Seyama, Sugimoto and Miyagi used organic film-coated quartz crystal resonators (QCRs) to measure the concentration and the chemical composition of volatile compounds for indoor air monitoring. Principal component analysis (PCA) - a statistical technique - was performed to categorise 23 kinds of volatile compounds into three general classes: alkanes, alcohols and aromatic carbons, indicating the behaviour of the occupant (e.g. smoking or having cheese) and assisting in welfare treatment (e.g. a baby with a wet nappy) in an indoor environment.

In remote sensing, *image fusion* [24] plays a vital role in extracting useful information from multisource and multitemporal images which are increasing in quantity, due to the advancement of modern vision technology. Many techniques have been developed to exploit these data. For instance, Bruzzone, Prieto and Serpico [25] used a neural-statistical approach to classify remotely sensed images. Not only were the spatial fea-

11

tures from multiple sensors considered, but also the temporal correlation at each pixel was capitalised to achieve a better classification result. In addition to image fusion at pixel level, there are another three levels of fusion, namely signal, feature and symbol (see [26] for more details).

2.1.2 Sensor fusion mechanisms

In general, there are four types of sensor fusion mechanisms: estimation, classification, inference and artificial intelligence methods. Estimation methods use redundant information from multiple sensors at their lowest representation level of information (i.e. signal or pixel level) to predict the current state of the measurand(s) and to reduce the expected variance. Classification methods utilise information from multiple sensors at one level higher (i.e. feature level). Key features for dichotomy detection are extracted from the measurements and are subsequently compared with *priori* classes. Moving one level even higher, the inference or decision methods fuse measurements in the most abstractive manner. Symbols representing decisions are integrated to increase in truth or the probability values of events. Like the classification methods, the inference methods are used widely in pattern recognition applications. Artificial intelligence methods, unlike others, offer a unique way to integrate different measurements at all levels (ie. signal, pixel, feature and symbol). Moreover, they can be auto-calibrated against any gradual changes in the baselines of *priori* classes. Example algorithms available in each method will be described in brief in the following sections.

2.1.2.1 Estimation methods

The simplest method is to take a weighted average $Y(t)$ of the redundant measurements $\{S_1, S_2, S_3, ..., S_N\}$ from N number of sensors at a time t. The linear model can be defined as:

$$Y(t) = \sum_{n=1}^{N} h_n(t) S_n(t). \tag{2.1}$$

The parameter h_n is the weight for n^{th} sensor and $\sum_{n=1}^{N} h_n = 1$. Intuitively, h_n should be a large value when the nth sensor provides accurate information. If there is much noise in its measurements, a small value should be assigned to h_n.
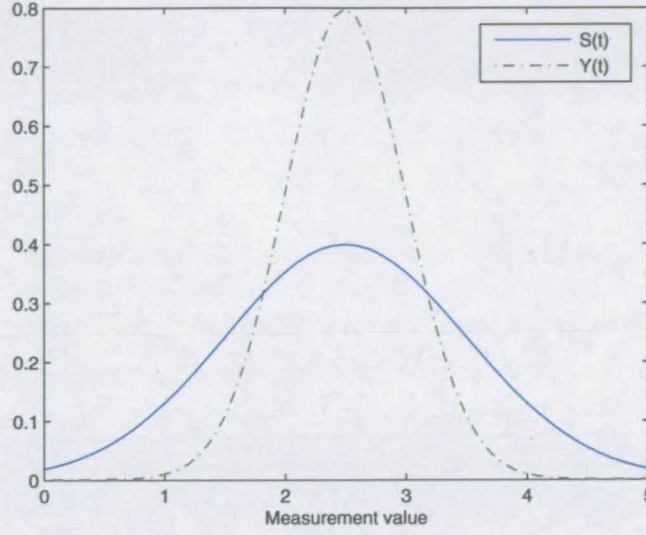
Figure 2.1: *Distribution plots for measurements taken from two sensors, where the standard deviation of S(t), σ_2 is greater than the standard deviation of Y(t), σ_1.*

Another method is the Kalman filter which is more optimized in a statistical sense than the weighted average. In this method, measurements are assumed to take the form of a white Gaussian distribution. The standard deviation of the distribution reflects the accuracy level of the measurements. If the measurements are found to be consistent, naturally the Gaussian distribution curve will be sharp with a high amplitude and a small variance. This means the measurements are more reliable. Otherwise, a broad-spread and low-amplitude distribution caused by noisy measurements, quite logically, indicates low confidence in these measurements.

Assuming that the measurand a of interest is changing at a constant speed, then it can be modelled as:

$$a(t) = A \times a(t-1) + B \times b(t) + w(t) \tag{2.2}$$

where A and B are two known transition matrices, $b(t)$ is an input vector, and $w(t)$ is a zero-mean random variable. Let $\Phi(t)$ be the transfer characteristic of a sensor, then its measurement can be modelled as:

$$S(t) = \Phi(t)a(t) + v(t) \tag{2.3}$$

where v(t) is the zero-mean random measurement noise. The estimate is written in the

following form [27, 28]:

$$Y(t+1) = \frac{\sigma_2^2(t)}{\sigma_1^2(t) + \sigma_2^2(t)} \cdot Y(t) + \frac{\sigma_1^2(t)}{\sigma_1^2(t) + \sigma_2^2(t)} \cdot S(t) \quad (2.4)$$

$$= Y(t) + K(t)[S(t) - Y(t)] \quad (2.5)$$

$$\text{with Kalman gain, } K(t) = \frac{\sigma_1^2(t)}{\sigma_1^2(t) + \sigma_2^2(t)} \quad (2.6)$$

$$\text{and} \quad \sigma_1^2(t+1) = A(t)\sigma_1^2(t)A^T(t) + Q(t) \quad (2.7)$$

where $\sigma_1^2(t)$, $\sigma_2^2(t)$ and $Q(t)$ represent the variances of $Y(t)$, $S(t)$ and $w(t)$ respectively. The latter two are assumed to be known. Eq.2.4 shows an interesting property of the estimate: if new measurement $S(t)$ becomes noisier (i.e. $\sigma_2 > \sigma_1$ as in Fig.2.1), the filter will automatically ameliorate the effect by reducing the contribution of $S(t)$ into the next estimate. Eq.2.5 illustrates that the next estimate $Y(t + 1)$ is optimized by minimizing the error, which is the difference between the actual measurement $S(t)$ and the current estimate $Y(t)$, also known as the "Innovation". When the measurand a has a nonlinear dynamic characteristic, Extended Kalman filter (EKF) can be used. In essence, the EKF provides an extra linearisation of a at a stable operating point.

2.1.2.2 Classification methods

Parametric template-matching method is widely used in image recognition. This method [29] searches for an object by scanning through the entire image pixel by pixel with a template, and calculating a normalised correlation value between the template and the portion of the search image at that position. The object is assumed to be at the position with the highest normalised correlation value. Conventionally, multiple base templates, also known as vertex templates, are used to improve the robustness against object deformation due to rotation, scaling, distortion and/or environmental changes (e.g. illumination variation). Each vertex template embeds one or more unique features. It is assumed that the object can be illustrated by a weighted sum of these carefully selected vertex templates. When one particular feature is dominant, its template parameter (i.e. weight) will be increased, and vice versa.

For non-imaging applications, linear discriminant analysis (LDA) is the most basic classification method [30]. It classifies an unknown sample X according to its distance d_x to the *priori* clusters. In the simpliest form, d_x refers to the distance between the sample
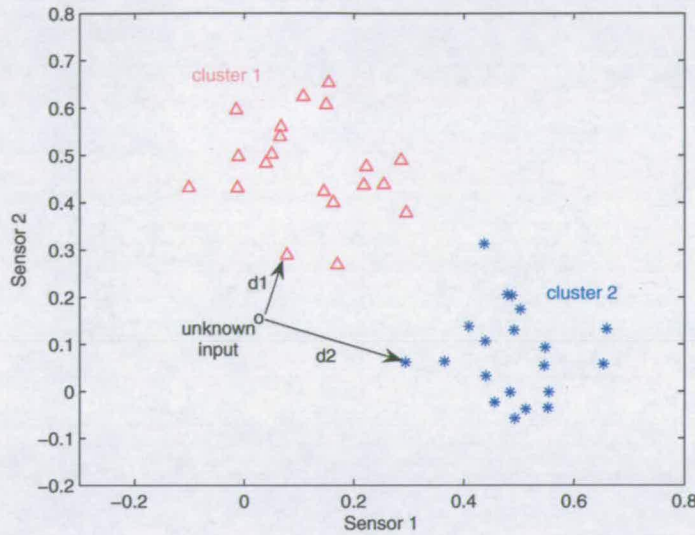
Figure 2.2: *The unknown input 'o' is classified according to its distances to the nearest samples from two priori clusters.*

X and the nearest sample from each *priori* cluster (see Fig.2.2). However, this simple discriminant method will misclassify when there is an odd sample from the wrong cluster located nearer to the sample X than the nearest sample from the correct cluster. A more robust method is to take several distance measurements for each cluster and compare their averages. This method is called K−nearest neighbour classification, where K is the number of distance measurements.

Another classification method based on distance measurements is cluster analysis. It partitions a dataset into subsets (clusters) such that the data in each subset shares some common traits. Cluster analysis can be hierarchical or partitional. With hierarchical algorithms, successive clusters are found using previously established clusters, whereas partitional algorithms determine all clusters in one go. Hierarchical algorithms can be agglomerative (bottom-up) or divisive (top-down). Agglomerative algorithms begin with each measurement as a separate cluster and merge them in successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters.

2.1.2.3 Inference methods

Bayesian inference is a multisensor data fusion method which is based on the rules of probability theory. Given an observation E, the Bayes' theorem calculates the posterior probability of a hypothesis H_0 with

$$P(H_0|E) = \frac{P(H_0)P(E|H_0)}{P(E)}.$$ (2.8)

The term $P(E)$ is called the marginal probability of E which is a normalising constant and can be calculated as the sum of all n mutually exclusive hypotheses $\sum_i^n P(E|H_i) \times P(H_i)$. The term $P(H_0)$ is called the *priori* probability of H_0 and is updated when new observational evidence is available. To fuse multiple observations from several disparate/ redundant sensors, their posterior probabilities are congruent multiplied to obtain a joint probability for each hypothesis. The hypothesis with the highest normalised joint probability will be regarded as the most probable answer.

The Bayesian method requires a good knowledge of all possible events. When there is a lack of such information, the inference result will understandably become less reliable. An alternative approach is to use the Dempster-Shafer theory of reasoning. It assigns probabilities to *sets* of possible subsets rather than a single subset as in the Bayesian method. *Belief functions* which model the probabilities of events take two values - *support* and *plausibility*, where *support* $\leq$ *plausibility*. The former indicates the amount of belief that supports the hypothesis directly while the latter sets the upper limit of the belief. Belief functions for different observations are integrated according to "Combination of Evidence" [31] to provide a coherent inference. When more supporting information becomes available, ignorance will diminish (i.e. *plausibility* $\simeq$ *support*) and the belief will become Bayesian.

A similar theory to Dempster-Shafer theory has been initiated by Lofti A. Zadeh and is called Fuzzy Logic. Each Fuzzy (sub)set is defined by a *membership function* to group different observations. Unlike crisp sets in classical mathematics, Fuzzy sets allow overlapping between different sets, where one observation can belong to two sets. Different observations from multiple sensors are integrated by simple logical operations such as *AND, OR* and *NEGATION*. The compositional result is then tranformed by *Defuzzification* into a single number indicating the most possible subset. More information about

Defuzzification can be found in [32].

2.1.2.4 Artificial Intelligence methods

In nature, biological systems are capable of sensing and performing pattern recognition effortlessly. For instance, a human being can sense immediately if the weather is cold or warm, and a fish can distinguish food from a stone in a deep dark sea. Thus, for many years, neuroscientists and statisticians have studied and proposed different forms of artificial neural network (ANN), in the hope of understanding how the biological systems actually work.

Typically, an ANN comprises layers of processing elements, known as neurons, which may be interconnected in various ways. The network is trained to store and recall the sensory information in response to different stimuli. The training algorithms are huge in variety and can be divided generally into *supervised* and *unsupervised* types. Supervised training has *priori* classes pre-defined and uses training rules such as "backpropagation" to search for a global solution to store the sensory information. ANNs that employ supervised training include Multi-Layer Perceptrons (MLP) and Radial Basis Functions (RBF). When *priori* classes are unavailable, unsupervised training algorithms such as "Minimizing Kullback-Leibler Divergence" are used. Examples of unsupervised ANNs include Kohonen's self-organizing map (SOM), Boltzmann Machine and Adaptive Resonance Theory (ART) [33].

Different sensor fusion mechanisms are performed at different representation levels of information [6]. Estimation methods have been successfully used for signal and pixel levels sensor fusion. Classification methods are used at feature level, while inference methods are for symbol level sensor fusion. In contrast, artificial intelligence (AI) methods can perform sensor fusion at all levels because they can be model-free and have a sufficient degree of freedom to fit complex non-linear relationships with adequate generalisation. Unsurprisingly, they are often considered as the advanced version of estimation, classification and inference methods.

2.2   ANNs in electrochemical sensor fusion

To date, the applications of ANNs are vast in variety, primarily in the fields of control and pattern recognition. In this section, however, only applications using electrochemical sensors, which are of relevance to this research work are discussed. The objective of this part of literature review is to examine related works and the latest research trend. Table 2.1 shows a list of example applications using ANNs to fuse electrochemical sensor signals.

There are two major sensing systems under this topic: the electronic nose and tongue. An electronic nose is defined as an instrument which comprises a sampling system, an array of chemical gas sensors with differing selectivity, and a computer with an appropriate pattern-classification algorithm, capable of analysing simple or complex gases, vapours or odours qualitatively [49]. A similar definition can be applied to an electronic tongue, except it analyses *liquid* samples. The primary aim of these two systems is to achieve similar sensing capability, if not better, to human's olfaction and taste systems. Inspired by nature, these systems are formed by an increasingly large array of sensors, thus local data fusion becomes more attractive and from necessity, akin to the situation in the *IDEAS* project.

Among the literature, Principal Component Analysis (PCA) is widely used to preprocess an array of sensor signals [36, 40–44, 48]. PCA is a linear transformation technique that can be used to select a new coordinate system for the dataset such that the greatest variance by any projection of the dataset comes to lie on the first axis (then called the first principal component, PC1), the second greatest variance on the second axis, and so on [50]. An example is illustrated in Fig.2.3. Because sensor signals are often not differential in a straight-forward way, PCA is used to extract the key features in the dataset. PCA is also used to reduce the data dimension when a large array of sensors (either identical or disparate) are used to improve robustness and/or to provide a more complete perspective of an entity. There is a similar statistical technique called "Discriminant Function Analysis (DFA)". Unlike PCA however, DFA allows non-orthogonal axial feature extraction hence provides better discrimination between clusters in classification problems [41].

One of the most popular ANNs is MLP mainly because of its strength in classification

No.	Application	Sensor	Algorithm	H/S[a]	Drift	Remark
1	Contaminant identification [34]	metal-oxide sensors	MLP	S	no	Use complimentary sensors (each with one ANN - associative memory/MLP)
2	Chemical sensing [35]	not given	Kohonen's SOM	S	yes	Adaptive model. Identify sampling rate to ensure trackability of drift
3	Juice classification [36]	metal-oxide and conductivity sensors	MLP	S	no	PCA to reduce input dimension
4	Gas identification [37]	metal-oxide sensors	SOM	S	yes	Adaptive model
5	Prediction of health of dairy cattle [38]	metal-oxide sensors	MLP	S	no	Time-dependent MLP
6	Breath alcohol detection [39]	metal-oxide sensors	MLP/RBF	S	no	Preprocessed with PCNN or rank-order-filter
7	Odorant classification [40]	conducting polymer sensors	GA-supervised MLP	S	no	Dimension reduction by PCA
8	Gas discrimination in an air-conditioned system [41]	metal-oxide sensors	PCA/DFA	S	no	Better result with DFA
9	Potato chips and cream classification [42]	conductivity and metal-oxide sensors	MLP/RBF	S	no	Use PCA to reduce input dimension
10	Classification of the strain and growth phase of cyanobacteria [43]	metal-oxide sensors	PCA/MLP/LVQ/ Fuzzy-ARTMAP	S	yes	Sensor signals are pre-processed Offline data fusion - not autocalibration
11	Water quality test [44]	conductivity sensors	PCA	S	no	
12	Detection of toxic gases [45]	metal-oxide sensors	MLP	S	no	
13	Online water pollution monitoring [46, 47]	ISFETs	Blind source separation	S	no	Plan to implement in MCU/DSP
14	Characterization of red wine [48]	Gas, optical and electro-chemical sensors	PCA	S	no	Fusion of three sensory systems

Table 2.1: *List of applications using ANN to fuse chemical sensor signals.*

[a]H/S refers to Hardware/Software implementation
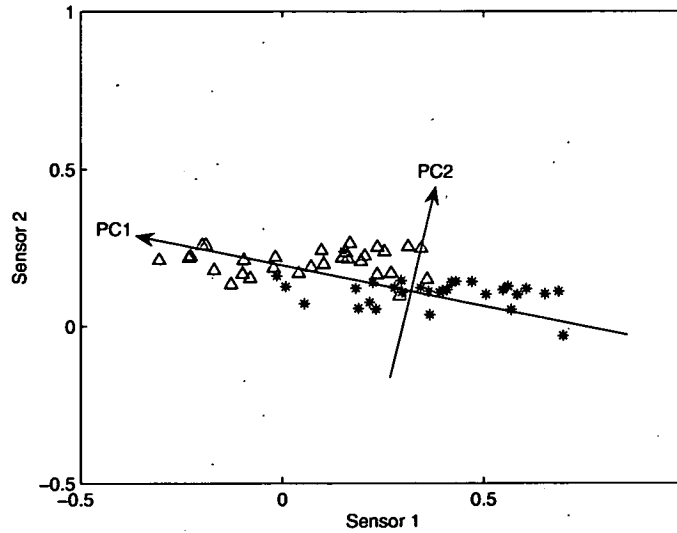
Figure 2.3: *An example of finding the first and second principal components for two over-lapping clusters. PC1 is perpendicular to PC2.*

and function approximation. It is in essence a static network, yet it can be adapted to process dynamic data by the addition of a tapped delay line, forming the Time Delay Neural Network [51]. MLP has been applied successfully in many real-world problems, ranging from contaminant identification [34] to juice classification [36], from dairy cattle's health prediction [38] to breath alcohol detection [39], and many more. Obviously, there are times when other types of ANN are preferred over MLP. One example is RBF [39, 42] which is known to be more efficient for interpolation and generalization of data than MLP, when the input data dimension is small. Another example is Kohonen's SOM [35] which learns without a teacher. This is particular by important when there is no pre-defined *priori* for clustering problems. Other applied ANNs include Learning Vector Quantization (LVQ) [43], Fuzzy-ARTMAP [43] and blind source separation [46–48].

Thus far, these electrochemical sensor fusions have merely been applied in software as it is a fairly straight-forward and rapid development task. Many commercial and non-commercial ANN simulators are now widely available[1], and most are accompanied with a good collection of established ANN algorithms. Even non-specialists in the ANN

---

[1]More details in URL: http://www.kcl.ac.uk/neuronet/products/.

field can perform the data fusion. However, such a software-based signal pre-processing scheme is often slow in computation and thus performed offline. This is not ideal when a real-time signal analysis is required on site. More importantly, it does not capitalize the parallelism in ANN, which, in nature, can offer high speed computation in pattern recognition applications.

With emerging applications such as distributed sensor networks, there is a growing interest to translate the ANN algorithms to hardware (microcontroller unit (MCU) and digital signal processor (DSP)) [46, 47]. Examples of implementation can be found commonly in robotic research works but are yet to be seen in electrochemical sensing. Translation to MCU and DSP is comparatively straight-forward due to their rich resources. However, it also means the hardware is not power- and silicon area-efficient. Therefore, researchers have started building custom hardware for the ANN algorithms by using analogue VLSI technology. Examples of such implementation will be reviewed in Section 2.4.

Interestingly, the drift issue has been raised on several occassions but is hardly ever addressed (refer to Table 2.1). In [43], Shin *et al.* employed offline processing to determine the growth phase of cyanobacteria in water for up to 40 days. Inevitably, sensor drift occurred but it was just treated as noise. No recalibration against drift was performed. The only attempts recorded in literature came from [35, 37], where adaptive Kohonen's SOMs were used to track a linear drift. Their simulation results highlighted the importance of annealing parameter K_σ to ensure the two classes were distinguishable on a trained SOM, despite the imposed drift. If K_σ was too small, the learning of SOM would cease learning at an early stage, probably even before the second class was presented. On the other hand, if K_σ was too large, the annealing would be too slow and the continuous learning might experience Catastrophic Interference (CI)[2]. The way they reduced the chance of a CI to occur was by feeding the two classes of data alternately into the SOM. Undoubtly, this drift tracking task would be much tougher if only a single class of data was available over a long period of time. Since the forementioned scenario is quite possible, if not typical, it will be considered in the drift tracking

---

[2]CI can be defined as the phenomenon that occurs when later training disrupts results of previous training and is characterized by the inability to incrementally learn sets of training patterns. CI is readily observed in studies of backpropagation. This phenomenon is also referred to as sequential learning and sometimes life long learning [52].

simulations in Chapter 6 later.

2.3   Counter-drift techniques

The reasons for drift are unknown dynamic process in the sensor system, e.g. poisoining and ageing of the sensor, or environmental changes, e.g. temperature and pressure variations. When sensor drift occurs, any subsequent data fusion will be impaired, if not impossible. Thus, the study of drift has been important to the sensor community. In this section, four general techniques to counteract drift are reviewed.

2.3.1   Recalibration

For each measurement, some form of uncertainty is encompassed inherently. There are two types of uncertainty, namely systematic errors (bias) and randomness. Repeating measurements can eliminate the latter but not the former. Change in bias while the environment is assumed to be constant is known as drift. One of the most popular techniques to counteract drift is to recalibrate either at pre-defined time intervals or when the measurements exceed the pre-defined range [53, 54]. In a more complex system such as [55], slow drift is distinguished from noise, sudden failure and changes in operating conditions by measuring not only the uncertainty of a measurement, but also the sum of uncertainty and the change of uncertainty for repeated measurements. Table 2.2 illustrates the characteristics of different sensor errors.

Traditionally, recalibrating electrochemical sensors is achieved with the use of a reference buffer [56]. But, such a technique is not suitable for miniature sensor systems. It is very difficult to scale-down the physical size of the buffer while ensuring the seal for the buffer is tight and no chemical reaction occurs [57]. An alternative recalibra-

Type of sensor error	Uncertainty	Sum of uncertainty	Change of uncertainty
Drift	not large	not small	small
Noise	small	about zero	large
Sudden failure	any	large	large
Changes in conditions	any	medium	medium

Table 2.2: *Characteristics of different sensor errors.*

tion technique is to apply on error correction algorithm. The sensors are continuously modelled using hidden variable models [58, 59] or Kohonen's SOM [37]. When the estimation error differs more than the permitted tolerance, the model parameters are updated to compensate for the drift.

2.3.2 Data filter

Another counteractive drift technique is to select the best estimation (i.e. one with the least drift) from a group of redundant sensors - i.e. a winner-takes-all approach [60]. Estimation of sensor signals is achieved with Kalman Filters. The array of sensors are divided into groups with overlaps. Each group is trained/modelled by a MLP. Output from the MLP which gives the least standard deviation (i.e. drift) is then presented as the best estimation. The standard deviation is additionally used as an indicator of the confidence in the estimations and hence the health of the sensor system. No drift-correction is attempted in this technique. So, the performance of the sensor system will degrade gradually since sensor drift is inevitable in all electrochemical sensors.

2.3.3 Drift insensitivity

The third technique is to make the sensor system output signal insensitive to drift. The most straight-forward way in achieving this is to perform measurements in *differential mode* [61, 62]. For instance, a pH sensing system employs one ion-sensitive FET (ISFET) and one ion-insensitive FET (REFET). The REFET uses buffered hydrogel or parylene as an ion-blocking membrane replacing the usual metal gate. Since both ISFET and REFET use a common reference electrode, they suffer the same amount of drift. The differential output has the drift effect cancelled out, and thus is independent of the drift. However, the two sensors suffer from "crosstalk" and hence the differential sensitivity is usually small.

Another way to make the sensor system output signal insensitive to drift is to model or predict drift based on historical data. When the reasons for drift are known, it is possible to develop mathematical models to compensate for the drift effect [63–65]. However, the causes of drift and their influence are normally unknown for electrochemical sensors. To resolve this, Sachenko *et al* [66] suggested using recurrent neural

network (RNN) to predict drift based on *short-term* historical data. In [67], a general mathematical model was employed instead to fit the drift pattern, and was optimised by minimising mean-square error (MSE) recursively.

Drift effect can also be eliminated by carefully removing it as a feature during data fusion. Artursson and Holmin [68, 69] employed PCA to extract key features from historical data, in which drift was identified as the first principal component. This component was then purposedly eliminated from contributing to the operating data fusion. Similarly, Lazzerini and Marcelloni [70] adopted a k-nearest neighbour (KNN) algorithm to extract features, and supervised fuzzy isodata (SFI) algorithm to select features to be fused. Assumed that a rich historical dataset is available, such a technique can permanently immunize the sensor system from drift effect.

2.3.4   Fault isolation

The fourth technique to counteract drift is to isolate drifting sensors from contributing to data fusion. Isolation is required when the standard deviation of the repeated measurements exceeds the permitted range. Drifting sensors are identified by comparing the actual sensor output and the estimated output of trained models such as autoassociative memory [71], fuzzy logic [72], MLP [73] and PCA [74]. Those with excessive estimation errors would be considered as faulty sensors and hence be isolated.

2.4   Neural Hardware in VLSI Technology

2.4.1   Introduction

In recent years, many neural networks have been realised in dedicated hardware to take advantage of the natural parallelism in ANN architecture. This parallelism is recognised as the enabling feature of our biological systems which allow us (homo sapiens) to perform more effective and faster pattern recognition, in spite of all the slow biological neurons, than a conventional von Neumann processor. This motivates engineers to develop neural hardware which can learn data structure and offer data fusion as in the biological systems.

Generally, there are two types of neural hardware. The first type tries to mimic biolog-

ical structures using very simple circuits - usually based on analogue VLSI technology. Meanwhile, the second type aims to support ANN as a mathematical tool and is commonly implemented in digital platforms (an architectural survey is available at [75]). Analogue neural hardware often has fixed functions but offers a simple, direct interface with the real world. Moreover, it can be achieved in compact layouts with a potentially low power requirement [76, 77]. On the other hand, digital neural hardware is more programmable, easier to communicate with other systems (especially for robust long-term storage and retrieval of ANN parameters in memory) and exhibits better immunity to noise and interference. More recently, hybrid techniques have also been exploited [78, 79] to gain the benefits from both technologies - analogue VLSI for data computation and digital VLSI for data communication and storage.

2.4.2 Supervised ANN-based hardware

Several successful implementations of supervised algorithms, mainly of MLP, have been reported in the literature [78–83]. The implementations are varied by their weight adaptation techniques. The three available techniques [84] are:

1. *off-chip learning*, where the neural hardware is not involved in the training process. The ANN parameters (known as weights) are trained using a computer with high precision. The trained weights are subsequently downloaded onto the neural hardware.

2. *chip-in-the-loop learning*, where the neural hardware is involved in feedforward propagation only. The calculation of new weights is done off-chip on a computer and is downloaded onto the neural hardware after each training cycle.

3. *on-chip learning*, where the entire training of the ANN is performed on-chip offering the possibility of adapting to the non-idealities of hardware (e.g. device-mismatch and non-linear response) and continuous training.

Early experiments [80, 82] show that the training of MLP requires at least 16 bits weight accuracy (which translates into a huge chip area). MLP with lower than 16 bits weight accuracy will not learn because the weight updates are often less than the quantisation step which prevents the weights from changing. One proposed solution is to use

a "weight perturbation [81]" algorithm where weights are constantly-updated at a pre-determined step size. Leong [79] reports that 6 bits weight accuracy is sufficient for the training of a MLP - a major improvement from previous MLP implementations with the more complex *backpropagation* [85] as theirs weight adaptation algorithm.

2.4.3   Unsupervised ANN-based hardware

Whilst the implementations of MLP are well established, the algorithm itself (MLP) is not adequate for applications where no *priori* classes are being predefined. A more suitable solution is to employ an unsupervised ANN, e.g. PCA, SOM, associative memory, Cellular Neural Network (CNN) and Boltzmann Machine. Examples of their implementations can be found in [86–91]. Among all the algorithms, CNN possesses the characteristics that suit analogue neural hardware implementation the most. It consists of simple generic circuitries as building blocks, which can be replicated many times. Moreover, it employs a local connectivity strategy to ease the connectivity complexity and also to permit very high speed operations (due to distributed computing).

Despite all these advantages, CNN like many others (e.g. PCA, SOM and AM) can only perform direct transformations (of possibly complex inputs to more "manageable" outputs) as in a look-up table. Each input is treated as a discrete variable. The underlying relationship (i.e. correlation) between the variables is not exploited as a feature in assisting the detection of dichotomy between two classes of measurements. This type of ANN is commonly known as a *deterministic* network. Its counterpart is a *probabilistic* network, also known as Bayesian belief network. It can model the joint probability distribution of all variables represented by each sensor node. Such capability makes the network more robust against incomplete input data entries (possibly caused by temporary or permanent occlusion of sensing surface), if compared with the deterministic ANNs. Other advantages offered uniquely by Bayesian belief networks are detailed in [92].

The *exact* inference methods in Bayesian belief networks can become intractable easily because their complexity grows exponentially with the number of variables [92–94]. Therefore, several *approximate* inference methods have been developed. Among them, there is one which is hardware-amenable, called the *Boltzmann Machine*. Example works that contribute to the VLSI implementation of Boltzmann Machine are [91, 95],

in which the networks are formed by stochastic binary neurons. The stochasticity is achieved by injecting artificially-generated noise (i.e. amplified on-chip thermal noise) into each neuron. These Boltzmann Machines have been tested successfully in the classic "XOR" problem, a standard benchmark to examine an ANN in a non-linear classification problem.

This type of Boltzmann Machine is however very restricted to binary problems. In [96], Chen and Murray demonstrated that binary neurons are not ideally suitable for modelling signals in continuous-valued format, a format in which most, if not all, real world signals appear in nature. In particular, the binary neurons are poor in modelling assymmetrical data distributions. The possible states for a binary neuron are '0' and '1'. Assume that a data distribution should be encoded by two hidden neurons with states {0,0}, {0,1} and {1,0}. However, simulation result revealed that there could also be data generated by state {1,1}, albeit at a low probability [96].

Furthermore, the Gibbs sampling method in the network optimization process is time-consuming. The network must be allowed to settle at equilibrium states before the weight updates can be calculated, which is a function of the Kullback-Leibler divergence for the data distributions of the initial and the equilibrium states. This leads to the proposal of Minimising Contrastive Divergence (MCD) by Hinton [97]. In the MCD learning rule, merely one-step Gibbs sampling is required for weight updates. This accelerates the training process.

In search for a Boltzmann Machine that can process real world signals directly, a new variant named "Continuous Restricted Boltzmann Machine (CRBM)" has been developed [96]. The MCD learning rule is adopted to ensure a faster convergence rate in the training process. Additionally, there is no connection between neurons in the same layer (visible/hidden), thus reducing the architecture into a simpler form. More importantly, the training rules are hardware amenable. Only simple mathematical operations, i.e. subtraction and multiplication, are used, which can be easily translated into hardware. The MCD learning rule has been implemented using analogue VLSI technology, and the neural hardware demonstrates its ability to perform on-chip learning to model simple distributions [98, 99]. Therefore, the CRBM is a natural choice of ANN for this thesis work.

2.5 Summary

In this chapter, the concept of intelligent sensor fusion with a formal definition, the motivations behind introducing ISF into a system and some examples of real world applications are presented initially. Then, the four types of sensor fusion mechanisms: estimation, classification, inference and artificial intelligence are briefly explained. In Section 2.2, research using ANNs for electrochemical sensing applications are reviewed. Most use an ANN as a mathematical tool in software and are not for the purpose of parallel computing. With the emergence of distributed sensor systems such as *WINS, SMART-DUST* and *picoNode,* it is anticipated that the demand for hardware-amenable ANNs will surge.

In Section 2.3, several counter-drift techniques have been discussed. They are recalibration, data filtering, drift insensitivity and fault isolation. In the *IDEAS* case, historical data is unlikely to be available and the number of sensors limited. These restrictions present recalibration as the only viable way to counter drift. In particular, unsupervised ANNs are of interest because they have the ability to adapt online, thus compensating for any experienced drift. In Section 2.4, the field of neural hardware is first introduced. Next, works contributing to the implementation of supervised and unsupervised ANNs with dedicated hardware are reviewed. Available options are narrowed down and eventually CRBM is identified as possibly the most ideal solution for the application. Later in Chapter 4, the CRBM will be examined in more detail and adapted into the application.

Chapter 3
Sensor models

There are two parts of preparatory work for the simulations in Chapters 5 and 6. This chapter discusses the first part which is the derivation of the sensor models. The second part is about the development of the neural system and will be presented in the next chapter (Chapter 4).

In this chapter, four different types of microelectronics sensors, designed and fabricated by Erik Johannessen (the sensor expert in the *IDEAS* team), are introduced. The main focus is on the sensor signal conditioning circuits, which are contributed by this thesis, because these circuits define the sensitivity and the operating range for the sensors. In this study, two generations of control chip design have been implemented onto silicon and tested. The subsequent sections explain the design considerations, circuit diagrams, physical layouts and test bench results. A discussion on the selection of the final sensor model concludes this chapter.

3.1   Integrated microelectronic sensors

3.1.1   Introduction

In the IDEAS project, four types of microelectronic sensor have been incorporated into the capsule in order to measure several physiological parameters, namely temperature, pH, dissolved oxygen (DO) and conductivity, in the GI tract. The temperature sensor measures the body core temperature and provides supplementary information for the temperature-dependent sensors. The pH sensor measures the acidity inside the stomach, which in turn can be used to identify the current location of the capsule since various portions of the digestive system have associated pH values. The dissolved oxygen sensor detects the activity of aerobic bacteria within the intestine, while the conductivity sensor gives an indirect estimation of the intestinal content by measuring the total level of dissolved solids (TDS).
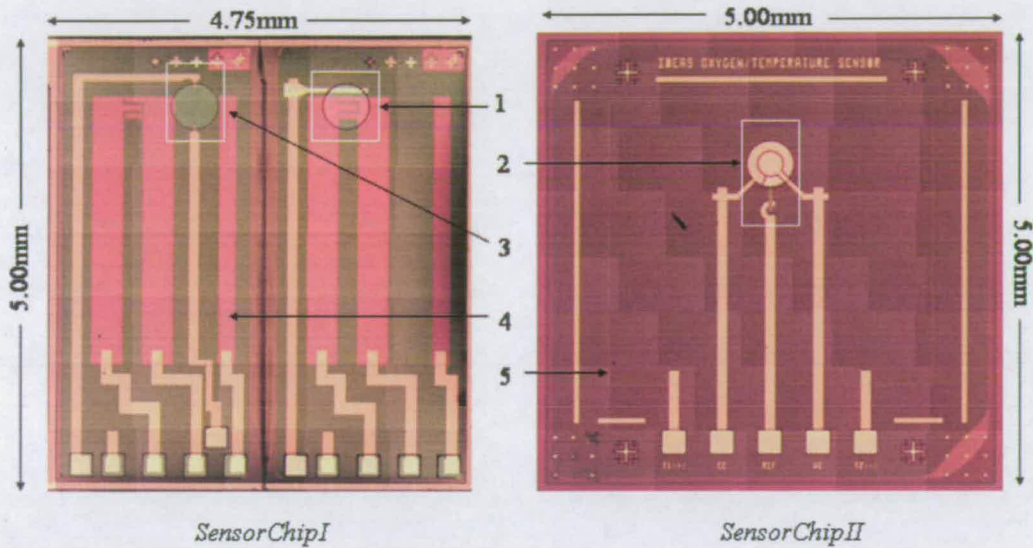
Figure 3.1: *Micrographs of the two sensor chips. SensorChipI comprises the temperature (4), pH (1) and conductivity (3) sensors whilst SensorChipII comprises the dissolved oxygen (2) and optional NiCr temperature (5) sensors.*

All four sensors were designed and fabricated on two silicon chips at Bioelectronics Research Centre, University of Glasgow: *SensorChipI* comprises the temperature, pH and conductivity sensors; *SensorChipII* houses the DO sensor and an optional nickel-chromium (NiCr) resistance thermometer. Fig.3.1 shows the micrographs of the two sensor chips. *SensorChipI* has a dimension of $5.00mm \times 4.75mm$ whilst *SensorChipII* is $5.00mm \times 5.00mm$. The following sections will state the specification of the sensors and explain briefly the basic operation of each sensor. The detailed information on the sensor fabrication process is reported in [100].

3.1.2 Temperature sensor

The main temperature sensor is a standard PN junction silicon diode, i.e. a narrow n-channel in a p-type bulk silicon (refer to Fig.3.1). It is based on a research product from École Supérieure D'Ingénieurs en Electronique et Electrotechnique (ESIEE), France.

The operation of the diode as a temperature sensor is defined by the standard equation

of the current flowing through a diode [101]:

$$I = I_s[exp^{\frac{qV}{kT}} - 1] \tag{3.1}$$

where

I_s = the reverse saturation current

q = electron charge ($1.6 \times 10^{-19} Coulombs$)

V = the voltage across the diode

k = the Boltzmann constant ($1.38 \times 10^{-23} JK^{-1}$)

T = the ambient temperature in $Kelvin$.

When the diode operates in the forward bias mode and $I \gg I_s$, the "-1" term can be dropped and Eq.3.1 can be re-arranged as:

$$V = \frac{kT}{q}(\ln I - \ln I_s). \tag{3.2}$$

This equation demonstrates that the voltage across the diode is linearly proportional to the ambient temperature, if I were held constant. Note that, this biasing current I scales the sensitivity of the temperature sensor. It is thus desired to be as large as possible in order to achieve high sensitivity, but also to have its upper boundary defined by the dynamic range of interest where the linear response lies.

3.1.3 pH sensor

The pH sensor is an ion-selective field effect transistor (ISFET) [102], again based on the ESIEE's product. Although ISFET is known to be less stable than the glass-based pH sensor [103], it has a much smaller footprint and more importantly, it can be fabricated using the standard CMOS process [104–108] - the enabling technology that fabricates redundant integrated circuit (IC) devices in batch and cost-effectively.

ISFET is based on the standard FET structure with an insulating layer in the gate region. When the ISFET is placed in solution, the solid/liquid interface potential E at the gate electrode, a function of the ion concentration to be determined, is measured according to the Nernst Equation:

$$E = E^0 - \frac{RT}{zF} \ln \left(\frac{a_{red}}{a_{ox}} \right). \tag{3.3}$$

E^0 denotes the standard electrode potential at $a_{red} = a_{ox}$ of the given redox couple, where a_{red} and a_{ox} refer to the chemical activities of the ions of interest which appear on the reduced and the oxidized sides of the electrode reaction respectively. On the other terms in Eq.3.3, R is the gas constant ($8.3143J/K - mol$), T the absolute temperature (in $Kelvin$), z the number of moles of electrons involved in the electrode reaction, and F the Faraday constant ($9.65 \times 10^4 C/mol$). For a pH-ISFET, its sensitivity defined by RT/F is 59.15 mV per unit pH at $T = 298.15K(25^oC)$, and its Nernst equation can be re-written as:

$$E = E^0 - 0.05915\ln(c_{H^+}) + 0.05915\ln(c_{ref}) \qquad (3.4)$$

where c_{H^+} represents the concentration of hydrogen ions in the solution and c_{ref} the concentration of ions in the reference electrolyte. Note that the redox process for a silver/silver chloride (Ag/AgCl) reference electrode is $AgCl + e^- \Longleftrightarrow Ag + Cl^-$, thus $z = 1$.

Fig.3.2(a) is a close-up view of the pH-ISFET, including a miniature on-chip Ag/AgCl reference electrode ($3 \times 10^{-2}mm^2$) and a floating gate ($15 \times 600\mu m^2$) covered with a 50nm thick proton sensitive layer of silicon nitride Si_3N_4 for pH detection. They are located in an electrolyte chamber (500μm-diameter, 70μm-depth) designed to contain 0.1M potassium chloride (KCl) gel electrolyte solution to promote a constant concentration of chloride ions. The chamber is defined in photocurable polyimide and is covered by a 50μm thick of cation-selective Nafion® membrane [109], forming a physical barrier between the chamber and the solution in which pH is to be measured. While the Nafion® membrane itself is very resist to decomposition, its seal (achieved with epoxy resin) is, as found later in the experiment, to be the source of sensor failure.

3.1.4 Dissolved oxygen sensor

The DO sensor is a three-electrode electrochemical cell [110]: the counter electrode (CE) is a 3-quarter, 200nm thick gold ring covering an area of $1 \times 10^{-1}mm^2$ (refer to Fig.3.2(b)); the working electrode (WE) comprises a microelectrode array of 57 circular gold electrodes (each 10μm in diameter), instead of single electrode with the same area size, to reduce response time due to diffusion effects [111]; the reference electrode (RE) is made of silver with an area of $1.5 \times 10^{-2}mm^2$ and a thickness of 500nm. The three electrodes are embedded in a 500μm diameter, 10nL electrolyte chamber defined in

(a)   (b)   (c)
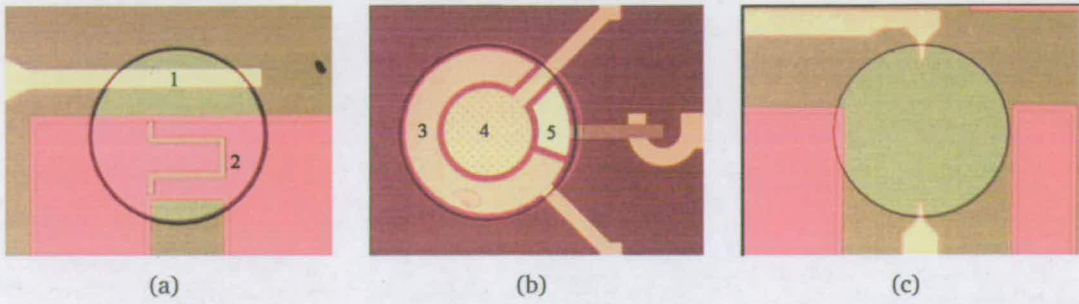
Figure 3.2: *The electrochemical sensors in close up views: (a) the pH sensor with a reference electrode (1) and a floating gate (2); (b) the dissolved oxygen sensor with a counter electrode (3), a working electrode (4) and a reference electrode (5); and (c) the dual-electrode conductivity sensor.*

polyimide similar to the ISFET's. The chamber is filled with 0.1M KCl gel electrolyte solution and is sealed with a 12μm thick, oxygen permeable teflon.

Fig.3.3(a) depicts an equivalent model of the three-electrode electrochemical cell [112]. Both R_s and R_{s2} denote the solution's resistance while R_{FC} and R_{FW} denote the Faradaic resistance of CE and WE respectively. C_C and C_W are the double layer capacitors[1] at CE and WE. Usually, R_{FC} and R_{s2} are relatively very small [114]. Therefore, at direct current (dc) level, the model can be simplified to a two-resistor model (refer to Fig.3.3(b)). If the constant bias voltage $v_{rw} = V_{WE} - V_{RE}$ reaches the redox potential of DO, the concentration of DO in the solution will be linearly proportional to the conductance of R_{FW}, thus the change in the current ($I_{rw} = v_{rw}/R_{FW}$) will reflect the change in the concentration of DO accordingly[2]. The control circuit that facilitates the electrochemical reaction is called a *potentiostat* [115].

3.1.5 Conductivity sensor

The conductivity sensor is a pair of direct contact, $5 \times 10^{-4} mm^2$ gold dual electrodes, spaced 400μm apart from each other. It operates according to the Ohm's Law ($V = IR$), where R is an inverse function of conductivity. To find out the conductivity of a solution,

---

[1]Double layer capacitor refers to the structure of charge accumulation and charge separation that always occurs at the interface when an electrode is immersed into an electrolyte solution. The excess charge on the electrode surface is compensated by an accumulation of excess ions of the opposite charge in the solution. The amount of charge is a function of the electrode potential [113].

[2]The input impedance at RE must be very large to avoid any leaking current into it.

(a) Equivalent DO model

(b) Simplified DO model at dc level
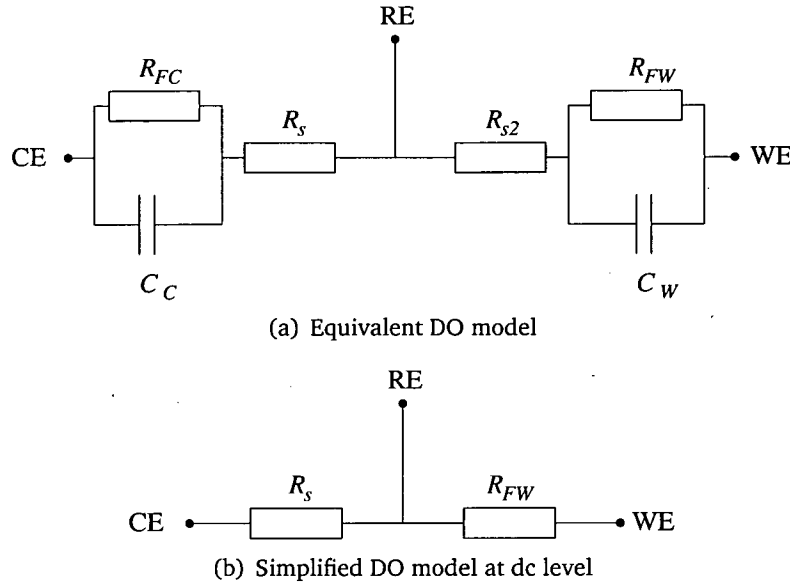
Figure 3.3: *Dissolved oxygen sensor models.*

a known voltage is applied across the dual electrodes. Commonly, alternating current (ac) type voltage source is used to minimise the effect of the faradaic resistance at the electrode surfaces (metal/liquid). The current flows between the electrodes, which is linearly proportional to conductivity, is then measured.

3.2 Sensor interface circuits

3.2.1 Introduction

The sensor interface circuits, also known as the signal conditioners, filter high frequency ($> 100kHz$) noise and amplify weak sensory signals. The output voltages from each sensor channel are then (analogue) multiplexed into single data stream before being presented to a 10-b analogue-to-digital converter (ADC) and the digital platform for further signal processing.

The signal conditioners form a major part of the control chip along with the digital platform and the optional RF circuits. There are two generations of design being fabricated, namely *IDEAS1* and *IDEAS2*, granting different circuit and system designs to be tested out: in *IDEAS1*, the signal conditioner designs are purposely kept as simple

as possible, the digital platform is merely a state machine with a low gate count and the radio frequency circuit is not implemented on chip; in *IDEAS2*, the digital platform is transformed into a simplified version of the Motorola 6805 central processing unit (CPU) and some RF circuits are introduced on chip.

Both the prototypes are fabricated by *austriamicrosystems* using a standard 3V, 0.6μm CMOS technology process with 3-metal and 2-poly layers, via the Europractise IC service. The Electronic Design Automation (EDA) tool, to design and simulate the analogue circuit designs, is the Cadence Virtuoso® Custom Design Platform, including the Schematic and Layout Editors, Spectre Circuit Simulator and Diva® physical verification tools.

The ADC, digital-to-analogue converter (DAC) and amplifiers used in the control chips are taken from the foundary cell library to save design effort and time. The amplifiers can operate in a "sleep" mode for power saving purpose. This facility is exploited in all the signal amplification stages, except the amplifiers which sense directly, because the setup time for certain electrochemical sensors is quite lengthy ($>$ 3mins).

3.2.2 First prototype: *IDEAS1*

3.2.2.1 Interface circuit design specifications

Generally, a targeted application defines the design specifications for the control chip, including the operating frequency, dynamic range and sensory sensitivity. For the *IDEAS* project, the microsystem is designed to be as generic as possible so that it can be applicable to many different biomedical and industrial applications, albeit with minor modifications. Therefore, the dynamic range is maximised such that it is as wide as possible whilst achieving an "acceptable" level of sensitivity. The following table lists the wishful front-end design specifications for the first generation control chip, *IDEAS1*.

3.2.2.2 System architecture's overview

IDEAS1 can be divided into mainly two modules, analogue and digital: the analogue module comprises the sensor interface circuits and the multiplexers; the digital module comprises a 10-b ADC, a 8-b DAC, a relaxation oscillator, a clock-divider and a state ma-

Type of sensor	Operating frequency (Hz)	Dynamic range	Sensitivity	Unit
Temperature	< 10	$0 - 70$	± 0.1	^{o}C
pH	< 10	$1 - 14$	± 0.5	pH
Dissolved oxygen	< 10	$0 - 10$	± 0.1	mgL^{-1}
Conductivity	< 10	$1 - 10$	± 0.1	$mScm^{-1}$

Table 3.1: *The front end design specifications for IDEAS1.*
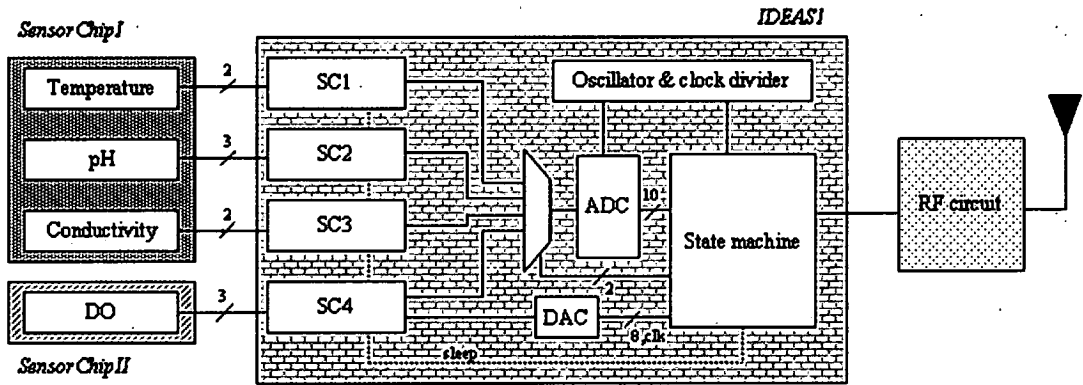
Figure 3.4: *Block diagram of IDEAS1 and its interfacing chips. SC refers to signal conditioner.*

chine. Fig.3.4 depicts an overview of the entire architecture, including their interfacing chips, namely *SensorChipI*, *SensorChipII* and the RF chip.

The choice of having *multiplexing* prior to analogue-to-digital conversion is made to keep the silicon area and power consumption to a minimal. Otherwise, the sensor channels will require one area- and power-expensive ADC each, particularly impractical in high-dimensional sensing applications. The DAC is used to search for the redox potential of the DO's potentiostat during the calibration phase. Once it is determined, the DAC's inputs are then fixed to either high or low during the capsule packaging stage, and hence the DAC sources a now constant input voltage to the DO signal conditioner.

In order to minimize the total number of off-chip components, an on-chip oscillator is deployed to generate a clock signal which has its frequency subsequently lowered and reshaped by the clock divider to become the global clock, t_{clk}. With this t_{clk}, the state

machine schedules the measurement steps. Additionally, the state machine also (a) performs some power management by sending a "sleep" signal across the chip to power down all the "idle" amplifiers, (b) packs the digitised datastream with some redundancy (to improve the wireless link's robustness), headers and footers (for recognition purpose at the receiving end), and (c) interfaces with the off-chip transmitter.

3.2.2.3   Design concept

The core design concept is to implement *simple* designs with minimal components, silicon area and power consumption, with minimal compromise in terms of sensory sensitivity. Such a method has very different priorities to standard circuit design practice, where the usual emphases are placed on the signal's stability and offset correction ahead of design complexity and power consumption. In this particular case, complicated circuitry, such as an instrumentation amplifier or a high order low pass filter, will not be used.

Furthermore, complex circuitry does not always promise the best solution. One good example is the pH-ISFET signal conditioning. The stability of an on-chip reference electrode poses a major challenge upon its integration with the pH sensor, because any instability in the reference electrode will result in undesirable sensor drift. One possible solution is to use differential measurement with a combination of one ISFET and one ion-insensitive REFET (reference field effect transistor). By default, the differential sensitivity is small. However, it can be enhanced by:

I. reducing the pH-sensitivity of REFET with a buffered hydrogel [116], an ion-blocking parylene [117] or a polyvinylchloride (PVC) membrane [62],

II. increasing the pH-selectivity of ISFET by using inorganic gate material such as Ta_2O_5 or Si_3N_4 [118] instead of SiO_2, and/or

III. using an immobilized enzyme on the gate structure of the FET as an additional layer to either select or block ions of interest [119].

Incidently, the differential mode of operation is subjected to cross-talk between the two transistors if the protons (by-product of the enzyme reaction) diffuse to the REFET,

resulting in a false signal. As a result, the single ISFET sensor with an integrated reference electrode is still the preferred choice, in this case, for its simplicity in the sensor interface circuit design.

3.2.2.4 Circuit designs

The control chip is powered by two SR44 Ag_2O batteries thus the power supplies available are Vdda (+1.55V), Gnd (0V) and Vssa (-1.55V). The following describes the circuit design for each sensor.

Temperature sensor There are two stages in this signal conditioner; sensing and amplification. In the first stage (refer to Fig.3.5(a)), a constant current is sinking through the temperature sensor and the resistor R_1. The change in the temperature is effectively affecting the closed-loop gain of the amplifier OP1 and its output voltage, which is subsequently amplified by OP2 according to the predefined dynamic range and resolution of the temperature sensor, with careful selection of the combination of the bias voltage V_{bias}, resistors R_2 and R_3.

pH sensor The constant potential across the resistor R_1 provides a constant current flowing into the ISFET, as depicted in Fig.3.5(b). The bias voltage V_{bias} is used to adjust the amount of current sinking. The gate voltage is held constant at Gnd as a reference point. The output voltage, or more precisely the potential across the drain and source electrodes, is left floating in response to the change in the ISFET's threshold voltage, a function of the concentration of H^+ ions in the sensing environment.

DO sensor As illustrated by Fig.3.5(c), the signal conditioner can be divided into three sections, namely the potentiostat, current-to-voltage conversion (I-V) and amplification. The bias voltage V_{bias} is employed to search for the optimal redox potential during the calibration stage and then set it constant at the counter electrode with respect to the reference electrode. A feedback path formed by OP2 and R_2 ensures the stability of the applied potential. The current flowing between the reference and working electrodes, a function of DO's concentration in the sensing environment/solution, will be detected by OP3 and converted into voltage format before being amplified by OP4. The amplification stage is necessary because

(a) Temperature sensor

(b) pH sensor

(c) DO sensor

(d) Conductivity sensor
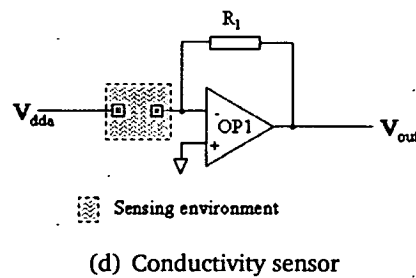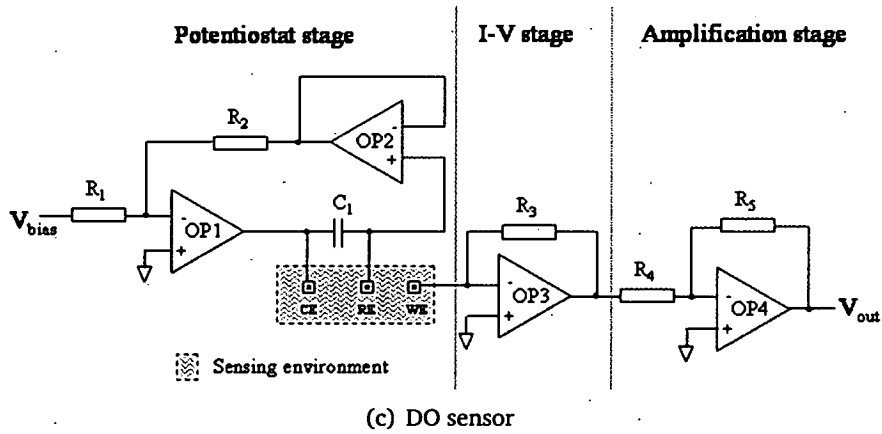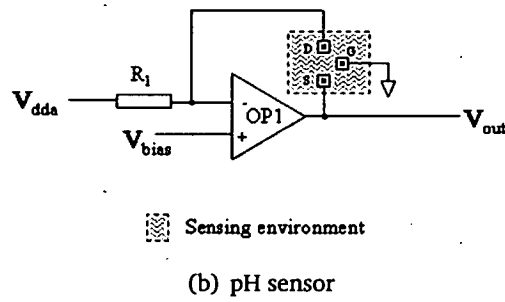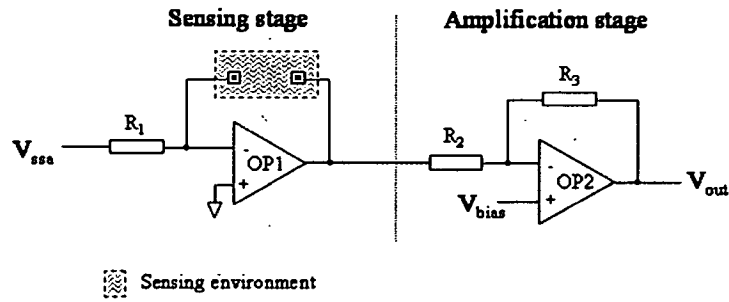
Figure 3.5: *Sensor conditioner designs.*

of the high impedance of the solution and the impracticality of implementing a huge resistor ($R_3 > 100k\Omega$) on silicon. The DO sensor's dynamic range of interest determines the selection of resistors R_4 and R_5.

Conductivity sensor To maximize the sensitivity of the sensor, the faradaic resistance at the electrode surfaces (metal/liquid) must be insignificant if compared with the solution's resistance. This is only achievable when the operating frequency is very high ($> 100kHz$) [120–122]. However, such a high frequency demands very high sampling rate[3] ($> 2MHz$) and power consumption to convert the sensory signal into digital datastream, therefore it is infeasible in this particular case. Instead, the conductivity will be measured at dc level with a simple setup, an inverting amplifier, as depicted by Fig.3.5(d). The sensor's dynamic range and sensitivity are balanced with the choice of resistor R_1.

3.2.2.5 Test bench results

The *IDEAS1* is fabricated and packaged into the standard 68-pin Ceramic Leadless Chip Carrier (CLCC68) chip format for test purposes, with 20 pads for power supplies, 44 for input/output and 4 unused (thus grounded). The chip is pad-limited and covers a silicon area of 4.5mm × 4.5mm. Its micrograph is shown in Fig.3.6 with an associated explanatory diagram.

The chip is tested by the author of this thesis on a laboratory bench with hardwired connections between it and the *SensorChipI*. Note that the *SensorChipII* is unavailable initially due to a severe electrolyte leakage problem with the on-chip reference electrode's chamber. The DO sensor is later tested by Erik Johannessen in oxygen saturated water with a standard laboratory potentiostat (Bioanalytical Systems, USA). The level of DO in the solution is monitored by a standard Clark O_2 electrode (Orion Research Inc, USA). The correctness in terms of functionality, of the DO signal conditioner, on the hand is verified separately.

During the testing, there is a constant 3mV peak-to-peak (equivalent to one least significant bit (LSB) of the ADC) noise present in the setup. Since the sensors in a harsh

---

[3]The ADC uses 11 clock cycles to convert an analogue signal into 10-b digital word and the Nyquist Sampling Theorem states that the sampling frequency must be greater than twice of the maximum frequency to avoid aliasing.

40
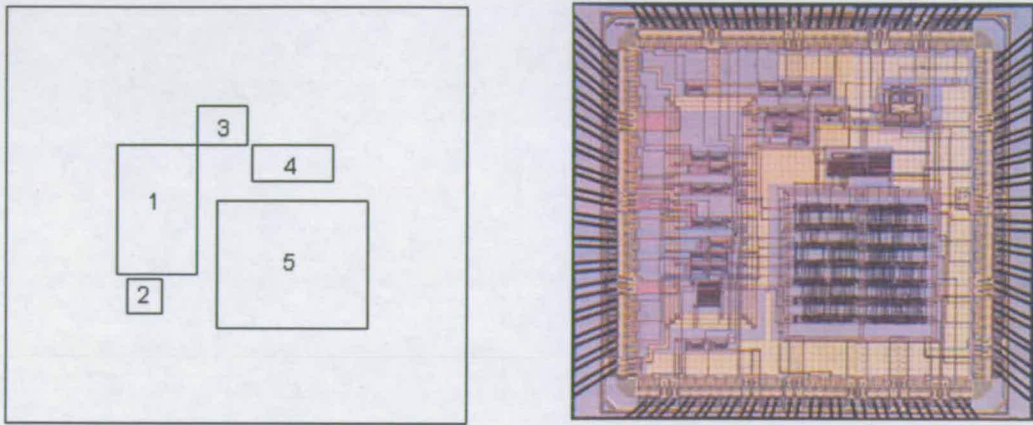
Figure 3.6: *The associated explanary diagram and the micrograph of the IDEAS1, comprising the sensor interface circuits (1), 8-b DAC (2), on-chip oscillator and clock divider (3), 10-b ADC (4) and state machine (5).*

environment will suffer larger noise, the second LSB is used for extra on noise margin. Hence, it is anticipated that the 10-b ADC has an effective resolution of 8-b. Fig.3.7(a)-(d) show the calibration results for temperature, pH, DO and conductivity sensors.

Temperature sensor The forward-biased silicon diode[4] operates with a constant current of $15\mu A$. The sensor responds linearly for a dynamic range from 0^oC to 70^oC with a sensitivity of $31.5mV/^oC$ after the signal amplification stage (refer to Fig.3.7(a)).

DO sensor Fig.3.7(b) depicts the linear response of the sensor within a dynamic range from 0 to 8mg/L. A sensitivity of $0.12mgL^{-1}nA^{-1}$ is achieved with an applied working electrode potential of -700mV with respect to the Ag/AgCl reference electrode.

pH sensor The pH-ISFET is calibrated with three standard buffers at pH 4, 7 and 10. As illustrated in Fig.3.7(c), the sensor exhibits a linear characteristic over a dynamic range from pH 4 to 10 with an average sensitivity of 23.4mV/pH.

Conductivity sensor The sensor is calibrated with several solid-state resistors, mimicking solutions with different conductivity levels. The signal conditioner, as expected, exhibits a logarithmic performance from 0.05 to 10mS/cm (Fig.3.7(d)).

---

[4]The sensor chip is immersed into a Phosphate Buffered Saline (PBS) tub.

(a) Temperature sensor

(b) DO sensor

(c) pH sensor

(d) Conductivity sensor
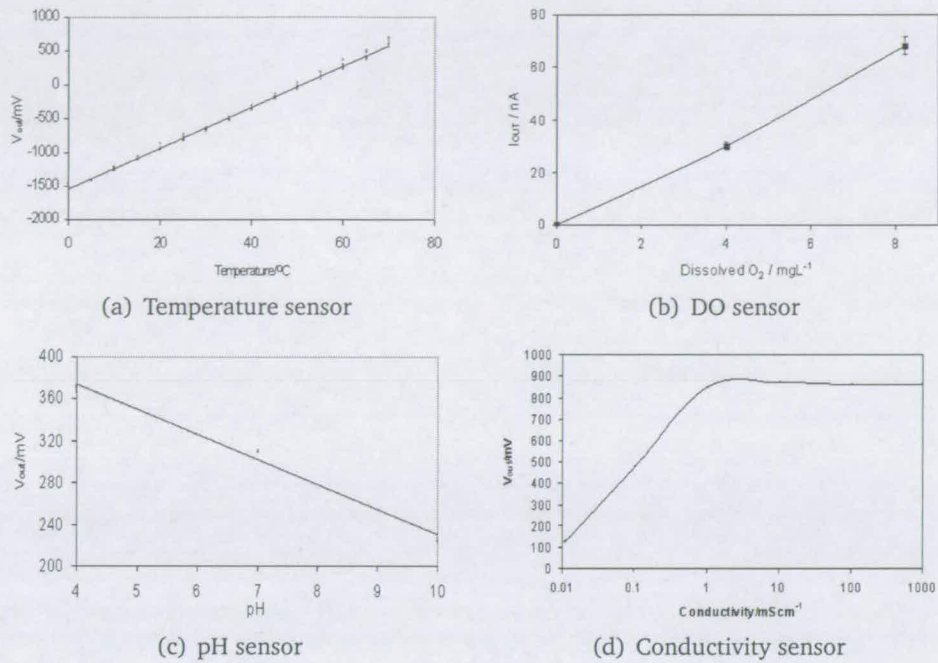
Figure 3.7: *Calibration results of the sensor conditioners based on five readings per calibration point.*

The calibration results are summarised by the following table in terms of the sensor dynamic range, resolution[5] and correlation coefficient[6] (R^2).

Type of sensor	Dynamic range	Resolution	Unit	R^2
Temperature	$0 - 70$	± 0.4	^{o}C	82.9%
pH	$4 - 10$	± 0.64	pH	99.3%
Dissolved oxygen	$0 - 8$	± 0.08	mgL^{-1}	95.0%
Conductivity	$0.05 - 10$	± 0.02	$mScm^{-1}$	99.5%

Table 3.2: *The summary of the IDEAS1 calibration results.*

3.2.2.6 Capsule packaging

The entire microsystem is eventually integrated into a biocompatible capsule which has a solid chemical resistant polyether-terketone (PEEK) coating. The microsystem includes the two sensor chips, the control chip *IDEAS1*, a 40MHz transmitter and two

---

[5]The resolution is derived from eight usable bits out of possible 10-b provided by the ADC.

[6]In this context, the correlation coefficient is a quantity which gives the quality of a least squares fitting to the measurements.
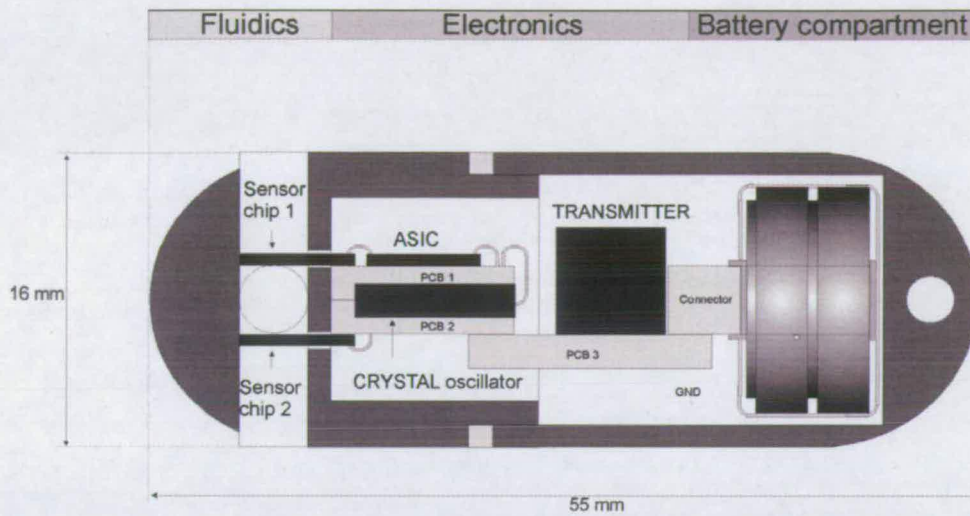
(a)
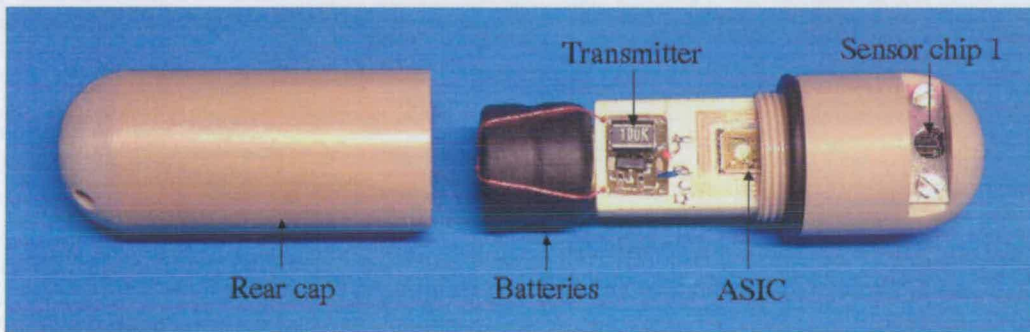
(b)

Figure 3.8: *Final packaged capsule in (a) schematic diagram and (b) photograph. This first prototype LIAP has its rear cap opened for the inner view of the microsystem.*

SR44 Ag_2O batteries. Its dimensions are 16mm $\times$ 55mm and its weight is 13.5g. Based on test bench results [100], the microsystem consumes 12.1mW of power in continuous mode (the worst case scenario) and the batteries last for 40 hours. Detailed experimental results on the capsule and the transmitter are reported in [100] and [123] respectively.

3.2.3   Second prototype: *IDEAS2*

3.2.3.1   Interface circuit design specifications

From the circuit design point of view, *IDEAS1* works as intended. Hence, the design specifications in terms of the operating frequency, dynamic range and resolution for the interface circuits in *IDEAS2* are almost identical to those in *IDEAS1*. Two key differences between the two generations of control chip are (a) four extra pH sensor channels are introduced as redundancy and (b) the resolution of the pH sensor is improved by sourcing more current into the ISFET in this second chip.

3.2.3.2   System architecture overview

One of the initial objectives in developing this second control chip is to further miniaturize the capsule packaging by exporing the integration of the analogue, digital and RF modules monolithically. In other words, the three modules need to be assembled together on the same silicon substrate without any bonding wires, thus reducing the total number of off-chip components and the packaging cost. Obviously, such integration gives raise to the concern of coupling noise between the noisy RF and the noise-sensitive analogue modules. To reduce the problem, different physical layout techniques are applied [124]. To name a few, there are the separation of low resistant power supply trees for each module, the implementation of on-chip decoupling capacitors and the introduction of guard rings around the noise-sensitive analogue module.

The second objective is to introduce an extra four pH sensory channels. This facilitates the investigation into feature extraction in a high dimensional space in a multisensor microsystem. Additionally, it also allows a potential improvement on the robustness of the microsystem by having redundancy. Ideally, the failure due to leakage of the reference electrolyte in each individual pH channel, the identified bottleneck of the microsystem, will now become less fatal.

The third objective is to replace the simple state machine in the digital platform with a modified Motorola 6805 microcontroller that enables the implementation of more signal processing functions on chip, at the expense of power consumption and silicon area. The fourth and final objective is to implement a direct-sequence spread spectrum (DS-SS) encoder which not only can improve the bit error rate (BER) of the wireless
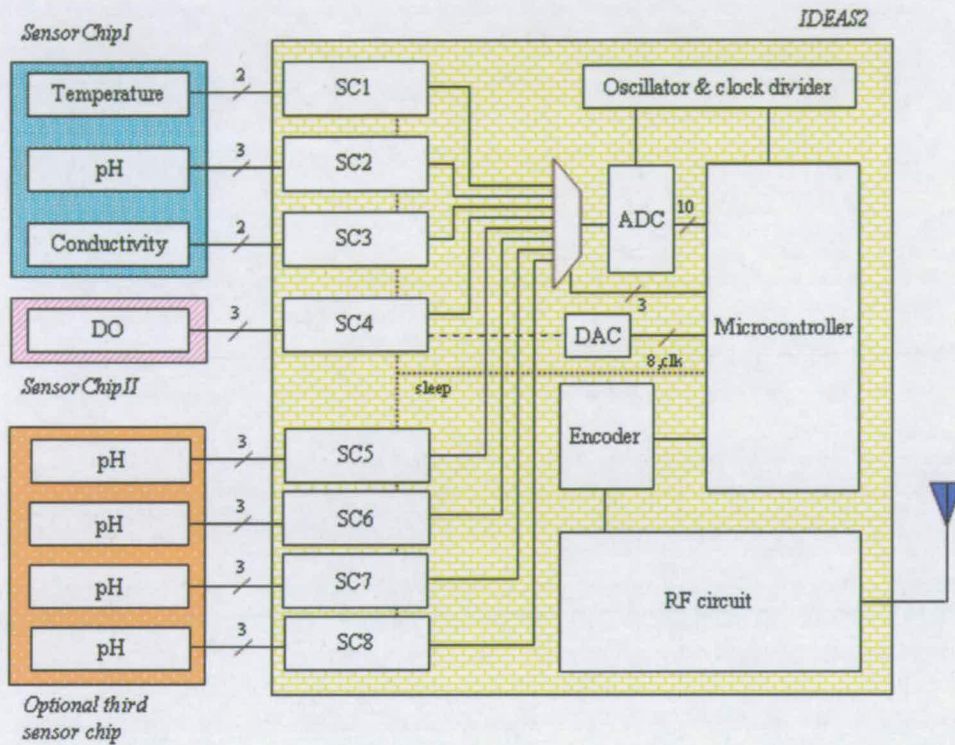
Figure 3.9: *Block diagram of IDEAS2 and its interfacing chips. SC refers to signal conditioner.*

data transmission [125], but also enable the use of multiple capsules with a single receiving base-station [126]. Each DS-SS encoder has a unique code sequence, thus the base-station can identify the sources of the received signals. Such coding system can be useful in applications such as waste water monitoring systems.

These four objectives define the specifications for this second generation chip *IDEAS2*. As depicted by Fig.3.9, *IDEAS2* interfaces with a third, optional sensor chip, which has four pH sensors, similar to the one in *SensorChipI*. An 8-to-1 multiplexer is introduced to fuse the eight sensory channel data into the ADC and the microcontroller. The digitised measurement datastream is subsequently fed into the DS-SS encoder. Its output is eventually passed to a transmitter circuit and an off-chip antenna.
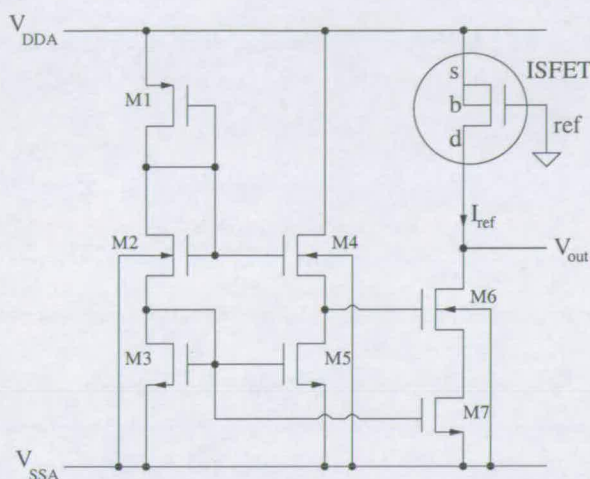
Figure 3.10: *Cascode current sink circuit for the pH sensor (ISFET). This scheme offers a high output resistance and a constant current over a broad range of output voltage [1].*

3.2.3.3 Circuit designs

As discussed in Section 3.2.3.1, the temperature, DO and conductivity sensor signal conditioners remain the same as in *IDEAS1*. However, a very different circuit is designed for the pH sensor (refer to Fig.3.10) where the ISFET forms as an active load to a cascode current sink. This type of current sink has the advantages of large output resistance and small non-saturation region in its *voltage-current* characteristic, resulting in a more constant current over a broader range of output voltage V_{out} [1]. With the reference electrode is tied to GND, the change in the environmental pH level swings V_{out} proportionally in a linear manner.

3.2.3.4 Test bench results

The sensor interfaces, digital platform and transmitter circuit are implemented on a 4.1mm $\times$ 4.1mm mixed signal SoC prototype, including a second ring of 100μm $\times$ 200μm pads added to facilitate the manual bonding for microsystem prototypes. The finished chips are returned from the foundary both as unpackaged dies (Fig.3.11) and in 84-pin J-Leaded Ceramic Chip Carrier (JLCC84) packages for test purposes, with 24 pads for power and 60 for I/O.

The packaged chip together with the pH and temperature sensors, is calibrated by the
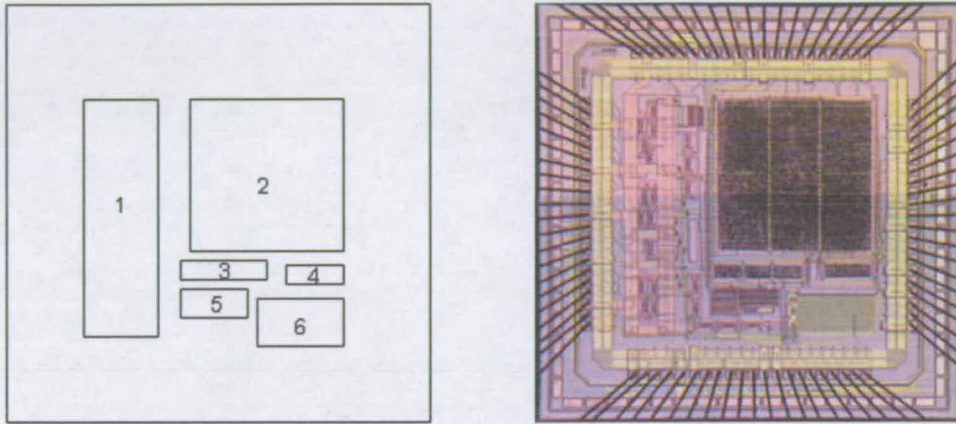
Figure 3.11: *The associated explanary diagram and the micrograph of the IDEAS2, comprising the sensor interface circuits (1), modified Motorola 6805 (2), clock divider (3), DS-SS coder (4), 10-b ADC (5) and RF circuit (6).*

author of this thesis with three pH buffers (pH 4, 7 and 10) on a standard test laboratory bench. The sink current I_{ref} (Fig.3.10) is measured to be $33\mu A$, which translates as an improvement in the pH sensor's sensitivity from 23.4mV/pH in *IDEAS1* to 43mV/pH in the new device. However, this also unveils a problem: the ISFETs have widespreading intrinsic threshold voltages ($0.28V < V_{th0} < 1.88V$) due to poor fabrication tolerance. Therefore, only a limited number of ISFET samples can respond linearly within the broad dynamic range predefined in the specifications of *IDEAS2*.

The unpackaged die is integrated as part of the capsule (Fig.3.8) and tested by the *IDEAS* team members as a complete system, including the wireless data link [127]. The capsule is immersed into a buffered solution with its pH level monitored by a standard pH meter. The capsule measures both the ambient temperature and pH values and transmits the data in bit stream to be picked up by a scanning receiver with a 1m range in radius. Upon its arrival, the data is stored immediately into the memory of a personal computer (PC). The measurements are then retrieved and displayed on the monitor using a custom graphical user interface (GUI) upon the user's request. Samples of the measurements are shown in Fig.3.12.

(a) Temperature measurement
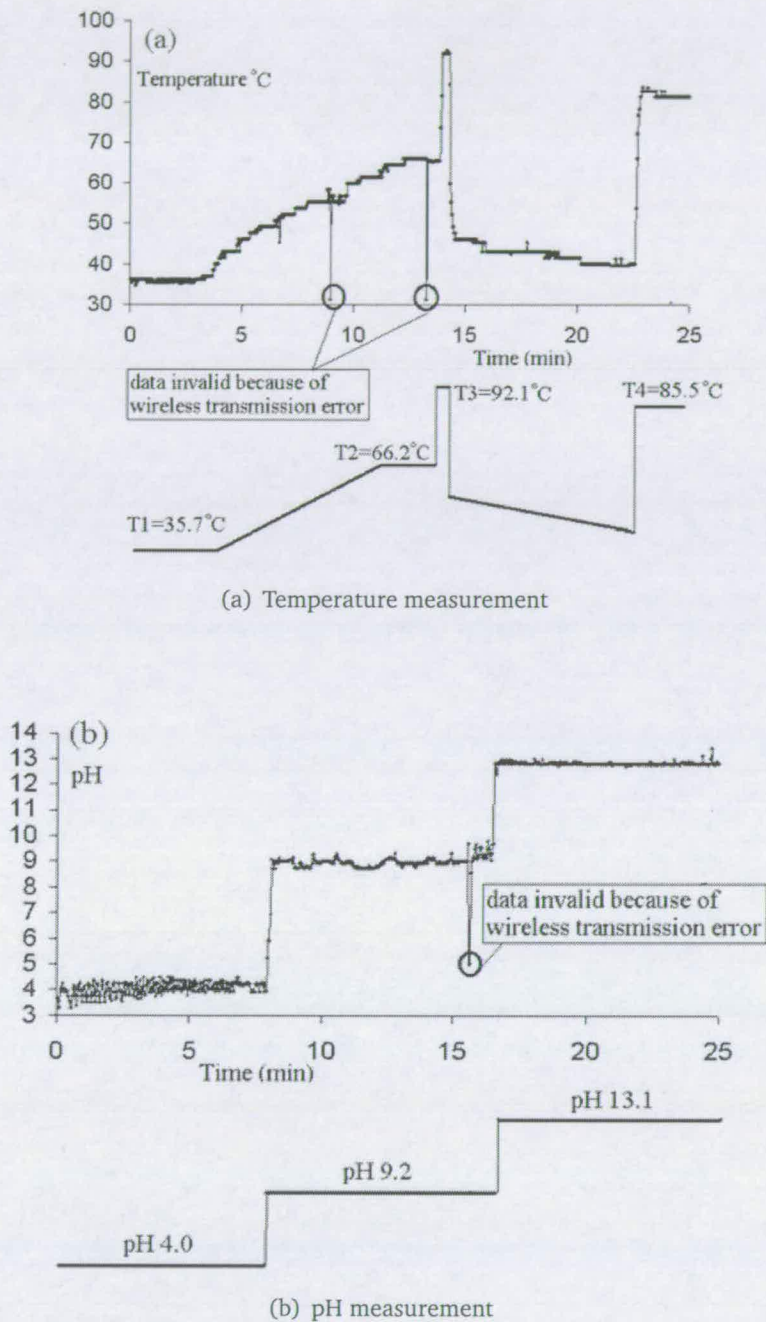
(b) pH measurement

Figure 3.12: *Measurements taken by the capsule over a period of 25 minutes. For the temperature measurements, the capsule was submerged in a 250ml glass bottle filled with reverse osmosis water at different temperatures. A thin wire K-type thermocouple (Radio Spares, U.K.) was used a reference. For the pH measurements, the capsule was submerged in different pH buffers and control measurements were performed with a standard lab pH electrode (Consort n.v., Belgium). These measurements were sent wirelessly by the capsule to an external receiver and a data recorder (i.e. a basestation). Graphs show the received signals.*

48

3.3 Discussion on sensor model selection

Actual sensory data are very rare and expensive to obtain. This may be a problem for training an artificial neural network. In this work, the calibration result of each sensor is mapped to a simple, linear model ($y = mx + c$). Such models can generated as much sensory data as required to achieve *good learning* in the ANN. More importantly, the artificial data allows the author to study and understand the neural system (in particular, the CRBM) in the early stage of this thesis work.

Thus far, four types of sensors have been introduced and reported with their test bench results on the two fabricated control chips, *IDEAS1* and *IDEAS2*. However, not all sensors provide satisfactory performance results. For instance, the conductivity sensor's performance is hampered by the double layer capacitance effect. One way to resolve the problem is to excite the sensor with a high frequency ac signal. Yet this requires a higher sampling rate in the ADC, which in turn consumes more power. This indirectly suggests that the conductivity sensor may not be suitable for implementation in a power-limited multisensor microsystem format, at least in this case.

Additionally, there is a stability problem in the DO sensor's on-chip reference electrode. Experimental results shows that it has a very short lifetime, typically two hours. This is unacceptable for the application - digestive tract monitoring system, which commonly requires the sensors to be functional over 8 hours. Another issue is found in the widely spread intrinsic threshold voltage of the pH sensors which restricts the amount of excitation current allowable for the sensors. Excessive current injection results in non-linear response in some pH sensors.

As a result, the feasible models are merely the temperature and pH sensors in *IDEAS1*. Calibration results from fabricated sensors [128] show that the temperature sensor's output voltage (mV) at time t can be expressed by:

$$V_{temp}(t) = 31.5 \times T(t) + T_0 + \tau_{temp}(t) + a_{temp} \cdot N_{temp}(t) \qquad (3.5)$$

where

$$
\begin{aligned}
T &= \text{ambient temperature } (^oC) \\
T_0 &= \text{-1505.10mV, output voltage at } 0^oC \\
\tau_{temp} &= \text{general non-linear drift function (mV)} \\
a_{temp} &= \text{background noise's magnitude (mV)} \\
N_{temp} &= \text{random Gaussian noise with unit magnitude and zero mean.}
\end{aligned}
$$

The term τ_{temp} is added for completeness but is often negligible in application. Similarly, the output voltage (mV) of the pH sensor can be modelled by:

$$
V_{pH}(t) = 18 \times T(t) - 23.4 \times pH(t) + V_0 + \tau_{pH}(t) + a_{pH} \cdot N_{pH}(t) \tag{3.6}
$$

where

$$
\begin{aligned}
pH &= \text{logarithmic concentration of } H^+ \text{ ions} \\
V_0 &= \text{offset voltage including the intrinsic threshold voltage of the ISFET (mV)} \\
\tau_{pH} &= \text{general non-linear drift function (mV)} \\
a_{pH} &= \text{background noise's magnitude (mV)} \\
N_{pH} &= \text{random Gaussian noise with unit magnitude and zero mean.}
\end{aligned}
$$

Note that the above equation includes a temperature term with a sensitivity of 18mV/^oC. The term τ_{pH} is negligible in the initial static neural training, but it has an important role in a dynamic environment (see Chapter 6).

3.4 Summary

Firstly, this chapter introduces the basic principles of four different microelectronics sensors, namely temperature, pH, DO and conductivity. The sensors are implemented[7] onto two silicon chips, *SensorChipI* and *SensorChipII*.

Subsequently, the design specifications for the sensor interface circuits are defined. The overview of system architecture and design concepts are also presented. In contrast to normal practice, design constraints (i.e. the silicon area and power consumption of the circuit) have higher priority than the sensor sensitivity itself. For instance, a simple current sink (see Section 3.2.3.3) is implemented to excite a pH sensor in place of a more commonly-used instrumentation amplifier. Note that an instrumentation

---

[7]The sensors are implemented by Erik Johannessen, University of Glasgow.

amplifier requires at least 3 amplifiers and multiple resistors to operate, thus its design opposes the design priorities defined in this particular work.

The sensor interface circuits are designed and integrated with the digital modules[8] and the transmitter[9] to form a control/communications chip. Two generations of control chips are fabricated by *austriamicrosystems* using a standard 0.6μm CMOS process technology. The chips are both tested on a standard laboratory bench and are used in the calibration of the sensors. Experimental results are reported and the circuits are working as intended.

Several problems have been identified in the sensors: the conductivity sensor requires more complex signal conditioning and measurand extraction; the DO sensor has a stability problem in its reference electrode; and the pH sensor has a relatively wide range of intrinsic threshold voltage for the 3V microsystems. Thus only the temperature sensor model and the pH sensor model with a lower sensitivity (23.4mV/pH) are usable in subsequent neural simulations. The models are extracted from their calibration readings, and each include a drift term. This term will come into play when data classification in a dynamic environment is examined in Chapter 6.

In short, the derivation of two simple but reasonably realistic sensor models (for subsequent simulations) has been presented.

---

[8]The digital module is designed and implemented by Lei Wang, University of Glasgow.
[9]The RF module is implemented by Mansour Ahmadian, The University of Edinburgh. It is only integrated on the same substrate in *IDEAS2*.

Chapter 4
Neural system

With the discussion of the sensor model made in Chapter 3, this chapter looks at the other key element in the preparatory work for subsequent simulations - the neural system. This system is formed by a Continuous Restricted Boltzmann Machine (CRBM) and a Single Layer Perceptron (SLP), as shown in Fig.4.1. The primary role for this system is to learn sensory data distributions and distinguish between two data clusters. One simple example of its possible applications in the *IDEAS* project is to provide information about the current location of the LIAP, because various parts of the digestive system have associated pH values.

This chapter provides a formal introduction to CRBM and a brief literature on SLP, in terms of their individual architectures and learning rules. To fit the neural system into the application, which suffers from sensor drift, some unique learning rules and methodology are necessary, and hence are proposed here. A discussion on the implication of such modifications in terms of hardware implementation is provided at the end of this chapter.

4.1 Architecture

4.1.1 System overview

The neural system operation can be viewed in two primary stages; unsupervised feature extraction and supervised (linear) classification. The aim is to render the feature-extraction stage adaptive and able to present a consistent set of features to the supervised classifer in the presence of drift and noise. This is a non-trivial task. When a microsystem is monitoring its surrounding environmental parameters, often there may be only a single class of data available over a long period of time. The long absence of another data class can cause a serious problem for a *learning* associative memory model because of catastrophic interference (CI). When only a *single* class of data is presented
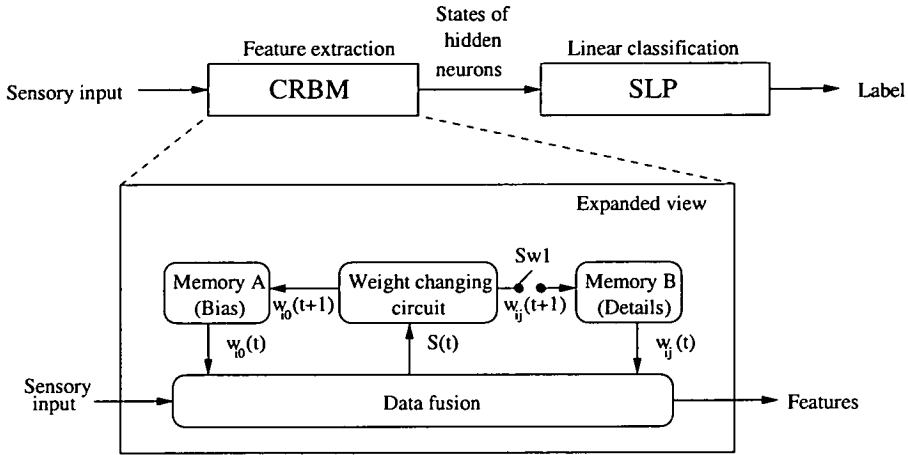
Figure 4.1: *An overview of the neural system and how the signal flows. Extracted features are represented by the states of the hidden neurons in the CRBM. The switch Sw1 is turned off after the initial learning stage (refer to Section 4.2.1).*

over a long period of time, the model perceives it as a completely new distribution. Therefore the model will adapt a new set of parameters to encode this "new" distribution and "forget" the existence of the previous distributions. Consequently the model loses its ability to classify.

The problem of CI came to light in the 1980's under the study of sudden breakdown in Hopfield network [129] and plasticity/stability dilemma in ART [130]. Connectionist network achieves remarkable abilities of generalization and graceful degradation in the presence of incomplete information by using a single set of weights as its memory. However, McCloskey & Cohen [131] and Ratcliff [132] showed that *using only a single set of weights* is also the root cause of the CI. Many attempts to solve the problem were initiated (see [133] for a review) but often required originally learnt patterns which, in the real world, were no longer available , and as such could not be added to the set of new patterns to be learnt. A major breakthrough occurred in 1995 when Robin [134] introduced a simple and elegant technique, named *pseudopatterns rehearsal*. The idea was to re-generate the originally learnt patterns with encoded weights from random input vectors and interleave these *pseudopatterns* with new patterns for the networks to learn. Such a technique is however computationally expensive and not ideal for power-limited operation.

Instead, a *constrained* learning approach is proposed here for this neural system. This

53

approach is motivated by the dual-memory models [135], simulating loosely the hippo-campal-neo-cortical separation (long- and short-memories). This approach exploits the fact that the weights of the hidden bias neuron w_{i0} encode the base line for the $(m-1)$-dimensional training data distributions while the weights of the hidden neurons w_{ij}, where $j = \{1, 2, ..., n-1\}$, describe the fine details of the distributions in a CRBM with an $m - n$ architecture [136]. The base line is a time-variant and shifts in proportion to the sensor drift. Therefore, it is possible to allocate a short-term memory for w_{i0} and a long-term (i.e. permanent) memory for the other weights. In this case, it is assumed the distributions have fixed pattern in terms of dispersion and shape. If the system can track the shift, then it is possible for it to compensate for the sensor drift. Section 4.2.3 will explain in details the unique learning rule and methodology that allow the CRBM to be adaptive and able to present consistent set of features to the supervised classifier.

4.1.2 Continuous Restricted Boltzmann Machine

4.1.2.1 General architecture

The CRBM is a generative model that is able to perform autonomous feature extraction and is based upon Hinton's Products-of-Experts architecture [137]. It has one visible and one hidden layer with only inter-layer connection. The visible and hidden neurons are connected by weight vectors {W}. Chen and Murray develop the model purposely so it can interface with analogue world directly. To do so, continuous valued neurons are used in contrary to the binary neurons in RBM [138]. Fig.4.2 shows a CRBM with three visible, four hidden and two (permanently 'on') bias neurons, v_0 and h_0.

4.1.2.2 A continuous stochastic neuron

Let s_i and s_j represent the states of a visible neuron i and a hidden neuron j respectively, and $w_{ij} = w_{ji}$ the bi-directional weights. The state-probabilities of the neurons are:

$$p(s_i = 1) = \frac{1}{1 + exp(-\sum_i w_{ij} s_j)} \tag{4.1}$$

$$\text{and} \quad p(s_j = 1) = \frac{1}{1 + exp(-\sum_j w_{ij} s_i)}, \tag{4.2}$$
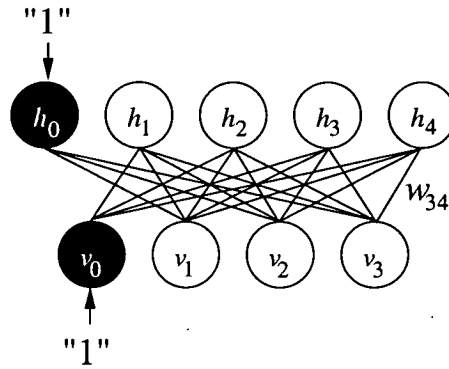
Figure 4.2: *A CRBM with three visible, four hidden and two bias neurons. Each neuron except the bias neurons which are always "1" has a sigmoidal activation function where its state s_i is confined by $-1 \leq s_i \leq +1$. The visible layer is connected to the hidden layer via symmetrical weights, i.e. $w_{ij} = w_{ji}$. This graph shows an example weight w_{34} which connects visible neuron v_3 to hidden neuron h_4.*
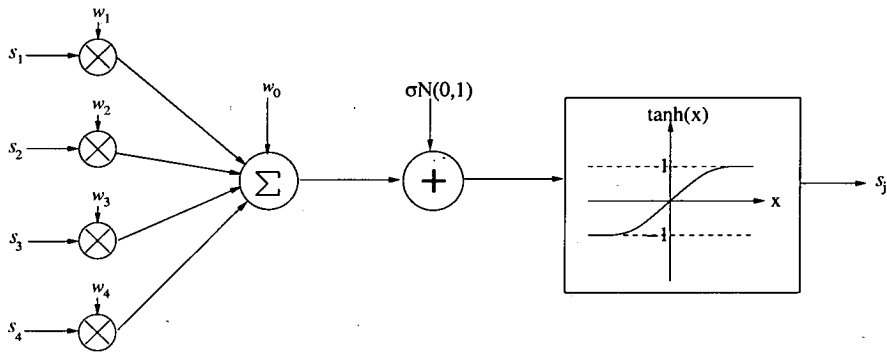
Figure 4.3: *A continuous stochastic neuron with four inputs.*

whereby the state of the neuron j is defined as:

$$s_j = \varphi_j \left(\sum_i w_{ij} s_i + \sigma \cdot N_j(0,1) \right) \tag{4.3}$$

$$\text{with} \quad \varphi_j(x_j) = \theta_L + (\theta_H - \theta_L) \cdot \frac{1}{1 + exp(-a_j x_j)} \tag{4.4}$$

where

σ = noise scaling constant : specific to each (visible or hidden) layer

$N_j(0,1)$ = sampled from unit-magnitude, zero-mean Gaussian noise source

a_j = noise control parameter : specific to each (visible or hidden) unit.

The product of σ and $N_j(0,1)$ forms a noise input component in the neuron, providing

it a sense of stochasticity. This is required in searching for the optimum model through the Gibbs sampling method. The product, $n_j = \sigma \cdot N_j(0,1)$ governed by the standard Gaussian distribution is represented by:

$$p(n_j) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(\frac{-n_j^2}{2\sigma^2}\right).$$

(4.5)

In other words, this continuous stochastic neuron is a Gaussian expert. Later in Chapter 5, the implications of this Gaussian nature as well as the magnitude of σ, on the CRBM learning ability in modelling both Gaussian and non-Gaussian datasets will be investigated.

The $\varphi_j(x)$ is a sigmoidal function with asymptotes at θ_L and θ_H. Commonly, it is set to be a *tanh* function (i.e. the asymptotes θ_L and θ_H are -1 and +1 respectively) so that the weight changes during the learning stage, can be in both positive and negative directions. Such setting allows the neural model in general to converge to the most optimal solution faster than uni-directional adaptation, where θ_L and θ_H are fixed at 0 and +1 or +2 respectively.

The parameter a_j controls the slope of the sigmoidal function, and thus the nature and the extent of the stochastic neuron behaviour. A small value of a_j leads to an almost-linear sigmoidal function, renders input noise [1] negligible and results a near-deterministic neuron. On the other hand, a large value of a_j leads towards a step function and results an approximately-binary stochastic neuron.

4.1.2.3 CRBM learning rule

The CRBM is a type of maximum likelihood model where its training involves modifying the parameters in the network to maximize the probability of the network generating the data as in the training set. In other words, the network must successfully model the probabilities of the training data. The CRBM employs "Minimizing Contrastive Divergence (MCD)" [97] as the learning rule for its weight, $\{w_{ij}\}$ and noise control parameter, $\{a_j\}$.

Before moving on to explain the cost function in the MCD rule, it is important to intro-

---

[1]This input noise is external, not the purposedly-injected n_j.

duce a quantity which measures the difference between two probability distributions p and q, called Kullback-Liebler divergence. It is defined as [139]:

$$
\begin{aligned}
KL(p,q) &= \sum_x p(x) \log \left(\frac{p(x)}{q(x)} \right) & (4.6) \\
&= -\sum_x p(x) \log(q(x)) + \sum_x p(x) \log(p(x)) & (4.7) \\
&= H(p,q) - H(p) & (4.8)
\end{aligned}
$$

where $H(p,q)$ denotes the cross-entropy of p and q, and $H(p)$ the entropy of p. As the cross entropy is always greater than or equal to the entropy, the Kullback-Leibler divergence is non-negative. If the two probability distributions p and q are identical, then $KL(p,q) = 0$.

The MCD rule aims to minimize *the difference* between $KL(Q^0, Q^\infty)$ and $KL(Q^1, Q^\infty)$, where Q^0 denotes the natural distribution over the visible variables at time zero for a Markov chain[2], Q^∞ the equilibrium distribution over the visible variable after prolonged Gibbs sampling[3], and Q^1 the distribution over one-step reconstruction of the data vectors that are generated by one full step of Gibbs sampling. The MCD rule differs from minimizing $KL(Q^0, Q^\infty)$, as practised in the log-likelihood maximization.

Hinton [97] argued that Q^1 is closer to Q^∞ than Q_0, thus $KL(Q^0, Q^\infty) \geq KL(Q^1, Q^\infty)$ and the convergence of the neural model is guaranteed. Since all transitions have non-zero probability in a Markov chain, when $Q^0 = Q^1$ (i.e. $Q^0 = Q^\infty$), the contrastive divergence will be zero and the neural model is perfect. Although the concept of MCD is purely based on an intuition, the results from several projects [137, 140–142] tend to support it.

---

[2] A Markov chain is a sequence of random values whose probabilities at a time interval depends upon the value of the number at the previous time.

[3] Gibbs sampling refers to drawing a sample from the posterior distribution of each variable given the current states of the other variables. With CRBM architecture (no intra-layer connection), the weights for all the experts can always be updated in parallel because they are conditionally independent.

The MCD update equations for w_{ij} and a_j are:

$$\Delta w_{ij} = \eta_w \left(\langle s_i s_j \rangle_0 - \langle \hat{s}_i \hat{s}_j \rangle_1 \right) \qquad (4.9)$$

$$\Delta a_j = \frac{\eta_a}{a_j^2} \left(\langle s_j^2 \rangle_0 - \langle \hat{s}_j^2 \rangle_1 \right) \qquad (4.10)$$

where

$\langle x \rangle_0$ = expectation value of x at time zero in a Markov chain

$\langle x \rangle_1$ = expectation value of x after one-step Gibb sampling

η_w = learning rate for weights, w_{ij}

η_a = learning rate for noise control parameter, a_j

The training procedures will be explained in Section 4.2.1.

4.1.3 Single Layer Perceptron

In 1943, McCulloch and Pitts [143] introduced the first neural model, Single Layer Perceptron (SLP). A SLP sums up weighted inputs and compares with a threshold, θ or $-w_0$, as illustrated by Fig.4.4. Let x_i be the input where $i = \{1, 2, 3, ..., n-1\}$, the output of the neural model y can be expressed mathematically as:

$$y = f_h \left[\sum_{x=1}^{n-1} (w_i x_i) - \theta \right] = f_h \left[\sum_{x=0}^{n-1} (w_i x_i) \right] \qquad (4.11)$$

where f_h is a Heaviside function,

$$f_h(x) = \begin{cases} 0 & : \quad x < 0 \\ 1 & : \quad x \geq 0 \end{cases}$$

The output function only produces a 0 or 1, denoting whether the neuron is on or off respectively.

Each SLP forms a linear discriminating function for two classes of data and is widely used as a linear classifier. To illustrate an example, let $y = 0$ for class A and $y = 1$ for

58
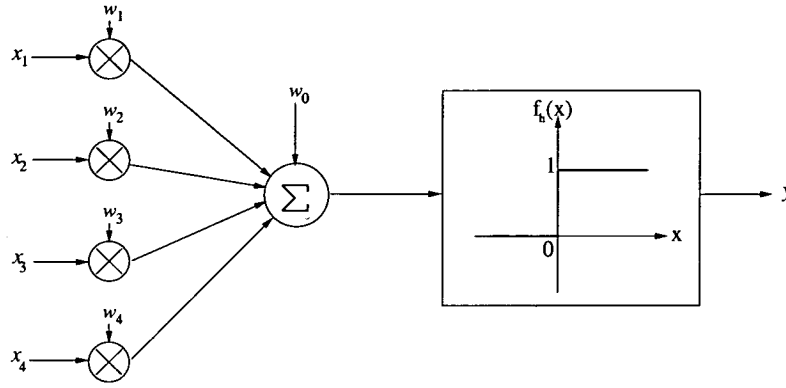
Figure 4.4: *A single layer perceptron with four inputs.*

class B in a two-dimensional space $\{x_1, x_2\}$. Thus, for class B data in Eq.4.11,

$$w_0 + w_1 x_1 + w_2 x_2 \geq 0$$

$$w_2 x_2 \geq -w_0 - w_1 x_1$$

$$x_2 \geq -\left(\frac{w_0}{w_2}\right) - \left(\frac{w_1}{w_2}\right) x_1$$

as illustrated in Fig.4.5. Comparing this to the equation of a straight line ($y = mx + c$), the slope of the line is controlled by w_1/w_2 and the intercept at x_2-axis is $-w_0/w_2$. Therefore, a SLP can discriminate linearly-separable data given a correct set of weights $\{w_0, w_1, w_2\}$.

Next, the learning rule for the SLP, the Widrow-Hoff delta rule [144], is introduced. It is modified from the basic Perceptron rule [145]. The Perceptron rule adjusts its weights by its input values directly. On the other hand, the delta rule calculates an error term Δ, the difference between the desired output and the weighted sum (i.e. the actual output). At each learning epoch, the weights are updated as follows:

$$w_{ij}(t + 1) = w_{ij}(t) + \eta \Delta \qquad (4.12)$$

with
$$\Delta = d(t) - y(t)$$

$$d(t) = \begin{cases} 0 & : \quad \text{if input from class A} \\ 1 & : \quad \text{if input from class B} \end{cases}$$

where η denotes the learning rate, $d(t)$ the desired response at time t and $y(t)$ the actual output. The derivation of this learning algorithm and the explanation on the
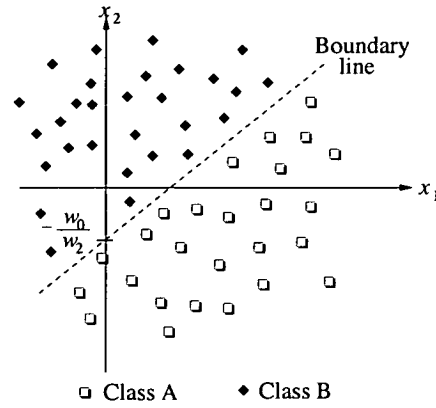
Figure 4.5: *Two classes of data separated by a linear boundary line that is defined by the weights of a SLP.*

convergence in learning can be found in [145].

4.2 Methodology

There are two preparatory steps to produce a trained neural system that can perform stochastic classification and auto-compensate any sensor drift. In the first step, the non-linear feature selector (CRBM) is trained, in an unsupervised manner, in order to learn the training data distributions. Upon completion, the linear classifier (SLP) is trained under supervision to map its inputs (selected features) with their correspond-ing labels $\{0, 1\}$. The neural system can then distinguish two classes of noisy sensory data. However, any drift in the data distributions will degrade the accuracy in data classification. To overcome this problem, the system must *learn* online to compensate the drift. Therefore this thesis proposes a unique method to do so in Section 4.2.3. The entire methodology flow can be summarised by Fig.4.6.

4.2.1 Training the CRBM

The CRBM is trained without supervision and in batch [4]. It has an architecture of $m : n$, where $(m - 1)$ is the dimension of training data and $(n - 1)$ is the number of hidden neurons. The number of features in the training data distributions determines

---

[4]Batch training only updates weights after all sets of inputs and their targets are presented. In contrast, online training allows immediate weight update after each set of inputs and targets is presented.
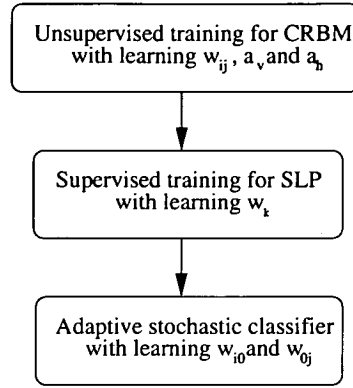
Figure 4.6: *Methodology to train the stochastic neural system.*

n. Although more hidden neurons offer the promise of a better model, Occam's Razor suggests that a parsimonious model is likely to be more useful and informative [146].

The training is evaluated by examining its 20-step Gibbs samping *reconstruction*. In this context, *reconstruction* refers to the ability of the model to generate distributions similar to the training data distributions, at the visible layer with its current parameters (weights) and disregards the initially-fed input data. The procedures to train the CRBM are:

1. Define w_{ij}, $i = \{1, 2, ..., m - 1\}$ and $j = \{1, 2, ..., n - 1\}$, to be the weight between visible neuron v_i and hidden neuron h_j, and the bias neurons v_0 and h_0 have a constant input of 1 each.

2. Determine empirically a suitable noise scaling constant σ for each layer.

3. Determine empirically suitable learning rates η_w and η_a.

4. Initialize w_{ij} by setting them to small random values.

5. Clamp the visibile neurons to the training data $\{X\}$ so $s_i = X$.

6. Infer s_j, the state of each hidden neuron with s_i by Eq.4.3.

7. Remove the clamp and let the visible neurons run freely.

8. Generate $\hat{s}_i$, a new state for each visible neuron, with s_j as inputs in Eq.4.3.

9. Re-infer $\hat{s}_j$, the corresponding state of each hidden neuron, with $\hat{s}_i$ as inputs in Eq.4.3.

10. Adapt the weight w_{ij} and the noise control paramter a with Eq.4.9 and 4.10 respectively.

11. Repeat steps 5-10 (full one-step Gibbs sampling) until the model can reconstruct the visible states similar to the training data.

4.2.2 Training the SLP

The SLP is trained under supervision and in batch. It takes the selected features $\{h_j\}$ as inputs, and outputs a label $\{0, 1\}$ with respect to the inputs at the visible neurons $\{v_i\}$. The mean square error (MSE), which is the average of the square of the difference between the desired output $d(t)$ and the actual output $y(t)$, is used as a performance indicator for the learning. The lower the MSE, the more accurate the data classification. The training procedures for the SLP are listed as follow:

1. Define $w_k(t)$, $k = \{1, 2, ..., n - 1\}$, to be the weight from input k at time t and the bias input, x_0 to be always 1.

2. Determine a suitable learning rate η_k by balancing between the number of learning epochs required for a convergence and the stability of the final SLP state. A large η_k moves the SLP faster towards the global minima but might oscillate at its final state. Conversely, a small η_k promises a more stable final state but requires longer learning epochs.

3. Initialize $w_k(0)$ by setting them to small random values.

4. Clamp the visible neurons of the CRBM to the training data $\{X(t)\}$ and the SLP to the desired response, $d(t)$.

5. Calculate the actual output $y(t)$ with Eq. 4.11.

6. Adapt $w_k(t + 1)$ with the Widrow-Hoff delta rule (refer to Eq. 4.12).

7. Repeat steps 4-6 for a pre-determined number of learning epochs or till a targeted MSE is achieved.

4.2.3 System in dynamic environment

Thus far, the discussion is on how to train a system with static data. Next, a novel methodology is proposed to allow the system to auto-calibrate itself and avoid CI in a dynamic environment. It achieves the aim by:

Task A allowing the CRBM to be adaptative to the latest distributions (i.e. the original distributions plus drift) and re-generating them at the visible layer, and

Task B ensuring that the CRBM continues to present *consistent* features to the SLP (linear classifier) in order to maintain high accuracy in the classification.

To accomplish **Task A**, the weight of the hidden bias neuron w_{i0} needs to be adaptive because it encodes the base line for the $(m-1)$-dimensional input data distributions. The learning rule for w_{i0} is as in Eq.4.9. The **Task B**, on the other hand, demands the weight of the visible bias neuron w_{0j} to be updated because it encodes the thresholds for the activities of the hidden neurons, which in turn determine the classification result. Incorrect threshold will result misclassification. Simply by updating these thresholds, consistent features can be presented to the non-adaptive SLP.

Based on Eq.4.3, it is obvious that an input drift (Δs_i) will result a corresponding change in the hidden neuron state (s_j), moving away from the known features to the SLP. Let $\Delta w_{0j}(t)$ be the amount of updating required and $\Delta s_i(t)$ denotes the average sensor drift in visible neuron v_i at time t. To maintain consistent features as inputs to the SLP,

$$\Delta w_{0j}(t) = -\left(\sum_{i=1}^{m-1} w_{ij}^* \Delta s_i(t)\right) \qquad (4.13)$$

where w_{ij}^* denotes the weight after the initial CRBM training session in static environment. Given that the model manages to trace the latest distributions and hence succeed in **Task A**, the drift can be expressed as:

$$\Delta s_i(t) = w_{i0}^* - w_{i0}(t). \qquad (4.14)$$

Substituting Eq.4.14 into Eq.4.13, we have

$$\Delta w_{0j}(t) = -\eta_{on}\left(\sum_{i=1}^{m-1} w_{ij}^*[w_{i0}^* - w_{i0}(t)]\right). \qquad (4.15)$$

In the above equation, η_{on} has been introduced as a scaling term to prevent the weights of the model from getting into saturation, in the event of learning from an instantaneous and huge (in magnitude) spiking noise, where $0 < \eta_{on} \le 1$.

During operation in a dynamic environment, the system (strictly speaking, it is the CRBM) is trained without supervision and online. The model learns the latest distribution with the following procedures.

1. Present the input sensor data to the visible neurons as $v_i(t)$.

2. Infer $h_j(t)$ the corresponding state of each hidden neuron with $v_i(t)$ in Eq.4.3.

3. Generate $\hat{v}_i(t)$ of each visible neuron with $h_j(t)$ in Eq.4.3.

4. Infer $\hat{h}_j(t)$ of each hidden neuron with $\hat{v}_i(t)$ in Eq.4.3.

5. Calculate $\Delta w_{ij}(t)$ with Eq.4.9.

6. Calculate $\Delta w_{0j}(t)$ with Eq.4.15.

7. Update the weight $w_{ij}(t)$ with $\Delta w_{i0}(t)$ and $\Delta w_{0j}(t)$.

The noise control parameter a is held constant, as it is after the initial CRBM learning stage. The purpose is to prohibit the model from perceiving the input data (from single class) as a new distribution, subsequently adapting to the "new" distribution (single cluster) and hence eventually losing its ability to classify. Besides a, the SLP also does not learn in this mode of operation. It classifies unknown input data by using the weights $w_k(T_K)$, where T_k refers to the final SLP learning epoch. The classification procedures are listed as below.

1. Set the noise scaling constant σ for hidden layer to be zero.

2. Present the input sensor data $X(t)$ to the visible neurons as $v_i(t)$.

3. Infer $h_j(t)$ the corresponding state of each hidden neuron with $v_i(t)$ in Eq.4.3.

4. Calculate the output $y(t)$ with $h_j(t)$ in Eq. 4.11.

4.3 Discussion on practical issues

As described in the previous section, some unique training methods and learning rules are necessary to equip the neural system with an auto-calibration capability. In this section, some practical issues of such methods and rules, in terms of hardware implementation, are probed. Note that there will be no actual hardware implementation of the neural system in this work. The following discussion is provided to highlight several important issues, namely the feasibility, cost and potential problems in translating the proposed learning rule into hardware.

The first question for the discussion is: *"is this new updating rule (Eq.4.15) hardware-amenable?"*. Failing to comply to this basic requirement will void the effort on the development of the CRBM. To answer, it is necessary to look at the type of circuits required first. Fig.4.7 depicts a possible circuit mapped from the equation. V_i is the subtraction of w_{i0}^* and $w_{i0}(t)$ where $i = \{1, 2, 3, ..., m - 1\}$. The multipliers ($Mult$) are fed with voltages (V_i and w_{ij}^*) and produce an output current I_i. The *current accumulator* sums up all the currents and relays I_{sum} as $\Delta w_{0j}(t)$ for a weight update. The digital control serves as a scheduler for the weight updating, while the input ε adjusts the scaling term η_{no}.

All the modules, namely *Sub, Mult, current accumulator* and *weight updating circuit*, have been implemented and reported in [147]. With minor modifications, such as changing the input for *Sub* from current-based to voltage and vice versa for the *Mult*, the modules are adequate for this new updating rule. Hence, it is reasonable to suggest that the new equation is equally hardware-amenable.

Obviously, this auto-calibration feature comes at a cost in terms of computational power and silicon area. But, a more worrying concern is if the neural system can still adapt when the drift velocity is high. In this context, drift velocity refers as the ratio of the drift to the sampling time interval. If high drift velocity occurs, the model will have difficulties to trace the drift, with a restricted η_{no}, and will lose track eventually. To overcome this, one might want to increase η_{no} but this makes the model more sensitive to noise. Alternatively, the sampling rate for the sensor data can be increased to "reduce" the drift rate indirectly at the expense of more power consumption.
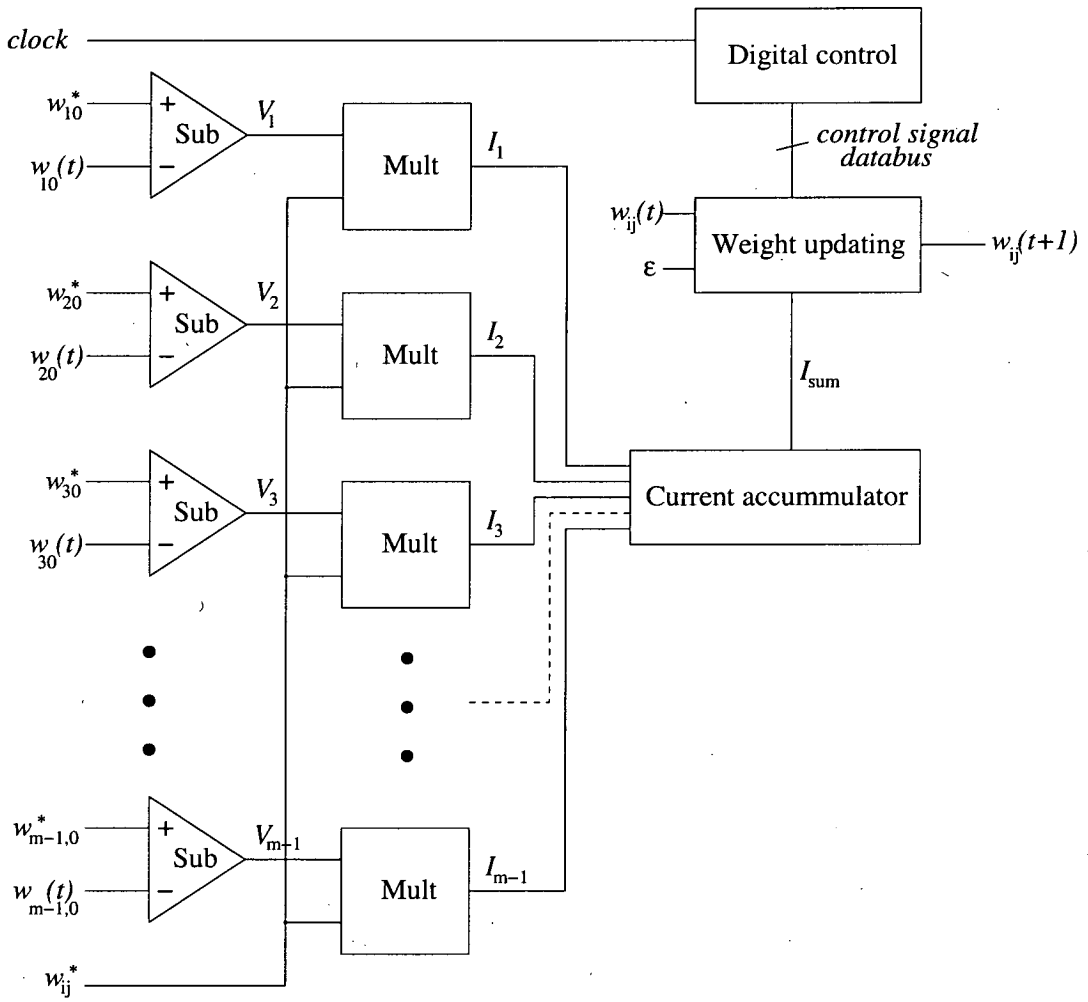
Figure 4.7: *The block diagram of the updating circuit for Eq.4.15.*

4.4 Summary

In this chapter, a neural system which is a combination of a CRBM and a SLP is first proposed. Both general architectures and learning rules of the neural model are introduced. The CRBM is developed initially by Chen and Murray from RBM to be able to interface with analogue signals directly. However, often there is only unbalanced, drifting and limited online data available in the application (the *IDEAS* project). Therefore, the CRBM is further developed in this thesis to facilitate an online learning process, with special concern on *Catastrophic Interference*. A new weight updating rule (Eq.4.15) is proposed, taking the advantage that the CRBM encodes the base line and fine details of the training data distributions differently and in separately controllable parameter subsets.

An unique methodology is outlined by this thesis to train the system *sequentially* to distinguish two classes of data, and to auto-compensate stochastic drift online. Finally, some practical issues (hardware-amenability, cost and potential problems) of the new learning rule (Eq.4.15) are discussed briefly. Based on the work in [147], the new rule is evidentially hardware-amenable and the the auto-calibration is feasible, at the expense of extra silicon area and power consumption.

Chapter 5
Classification simulation on sensor model and artificial data

The aim of this chapter is twofold: to explain the basic principles of this "neural" classification approach and to demonstrate that the proposed neural system is applicable to complex, meshed clusters problems. To date, CRBM has been examined with some toy data and electrocardiogram (ECG) data in [96]. This chapter wants to take one step further and examine it with much complex data in a systematic way. In the first part of this chapter, different features in the neural system will be identified in two- and multi-dimensional classification problems. The accuracy in classification with different levels of dichotomy between two training data clusters will also be studied. The second part of this chapter will focus on a more difficult learning problem - non-Gaussian clusters with meshed areas, examining with two different datasets. The performance of the classification with various numbers of hidden units and training epochs will then be studied to provide a guideline for system optimisation.

5.1 Classification trial one: simple, overlapping clusters

5.1.1 Two-dimensional (2D) clusters

The experiment involved one temperature sensor (S1), and one pH sensor (S2) with an intrinsic threshold voltage of 0.28V. The training data for the neural system (with four hidden units) was generated based on Eq.3.5 and 3.6 in Section 3.3. The conditions for the datasets were arbitrarily set to obtain two simple but overlapping clusters (Fig.5.1a), each with 200 samples in a Gaussian distribution.

The conditions set for class A data (output or label was 'HIGH') were 25°C (in temperature) and pH 4, while the conditions for class B data (output or label was 'LOW') were 40°C and pH 10. The CRBM weight w_{ij} had a learning rate of 1.0 while the other
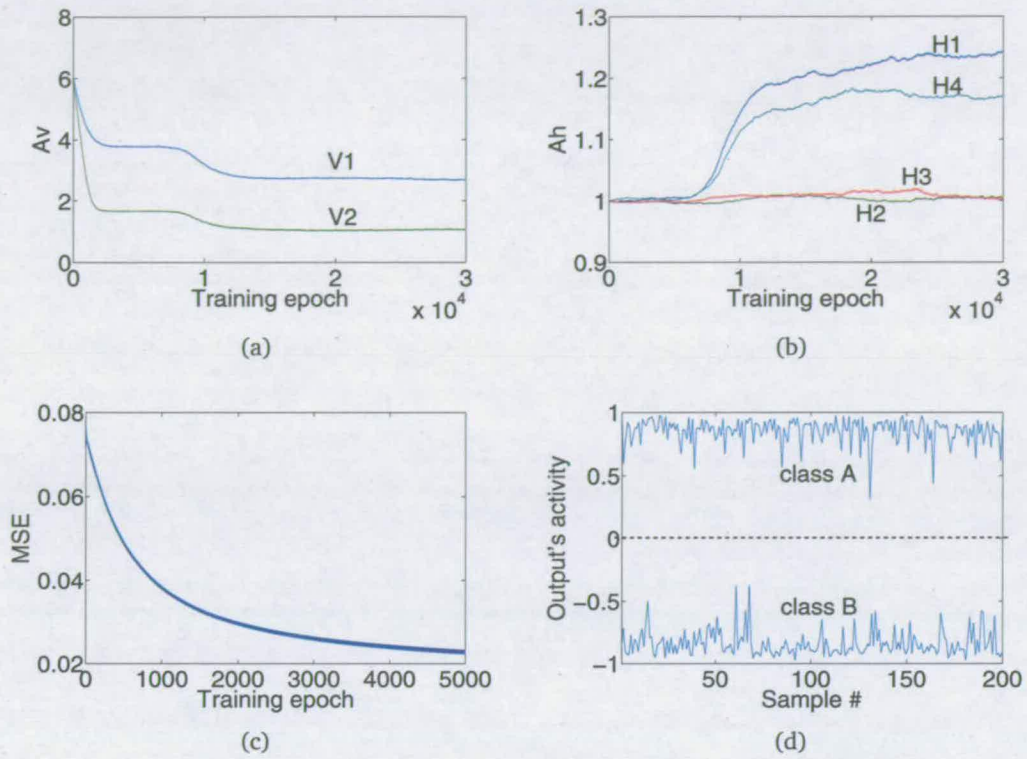
Figure 5.2: *Simulation results on 2D clusters: noise control parameter of visible layer A_v (a), noise control parameter of hidden layer A_h (b), MSE during the SLP training stage (c) and the SLP activity to 400 test samples (d).*

layer started to gain values, in particular the first and the fourth hidden units (H1 and H4), making them good detectors to distinguish the two data classes (Fig.5.2(b)). Fig.5.2(c) depicts the MSE measured in each training epoch for the SLP. Decreasing from the initial 0.0707 to the final 0.0223 indicates the SLP has learnt well. To prove it, a set of test data (200 samples for each class) was generated and fused through the learnt model. The corresponding activity of the SLP to this test data is presented in Fig.5.2(d). With the final w_k = {10.9061, -0.6689, -1.4227, 8.5075}, the difference in the two classes of data was exaggerated hence making the two clusters easily separable. Thresholding the output activity at zero sees 100% accuracy in the data classification by the neural system.
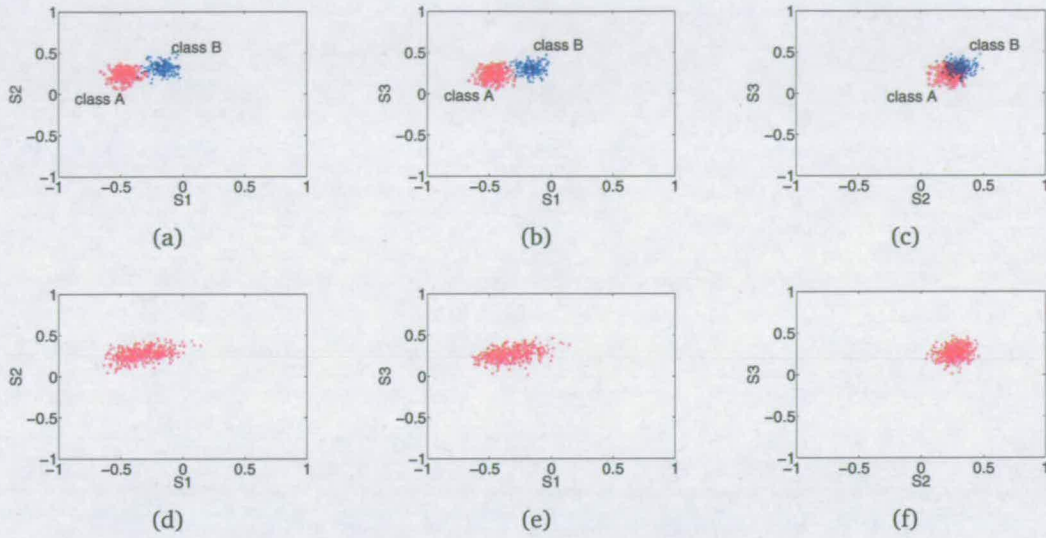
Figure 5.3: *Training data for the 11-dimensional clusters experiment: (a) temperature sensor (S1) and first pH sensor (S2); (b) temperature sensor (S1) and second pH sensor (S3); (c) first pH sensor (S2) and second pH sensor (S3). (d)-(f) are their corresponding 20-step Gibbs sampling reconstruction data.*

5.1.2 High-dimensional clusters

The experiment involved one temperature sensor (S1) and ten pH sensors (S2-11). The latter had an unique intrinsic threshold voltage each, V_{th0} = {0.50, 1.20, 0.28, 0.90, 1.60, 0.40, 1.10, 0.70, 1.88, 0.60}. Six hidden neurons were employed to capture the features in the training data distributions which were generated by using the sensor models in Section 3.3. Class A data was defined by the conditions where temperature was 25°C and pH 4, while the class B data was defined where temperature was 40°C and pH 10. The learning rates for the CRBM weight w_{ij} and the other parameters were 1.0 and 0.1 respectively. The neural system was trained with 200 samples from each class, according to the methodology (refer to Section 4.2). The numbers of training epoch were 30000 and 5000 for the CRBM and the SLP respectively.

Fig.5.3(a)-(c) depict the training data distributions in 2D while Fig.5.3(d)-(e) show their corresponding 20-step Gibbs sampling reconstruction data distributions. As depicted, the CRBM has encoded the training data distributions reasonably well in its weight and noise control parameters after 30000 training epochs. Further evidence was found in the noise control parameter A_v recorded over the training period (Fig.5.4(a)).

(a)

(b)

(c)

(d)
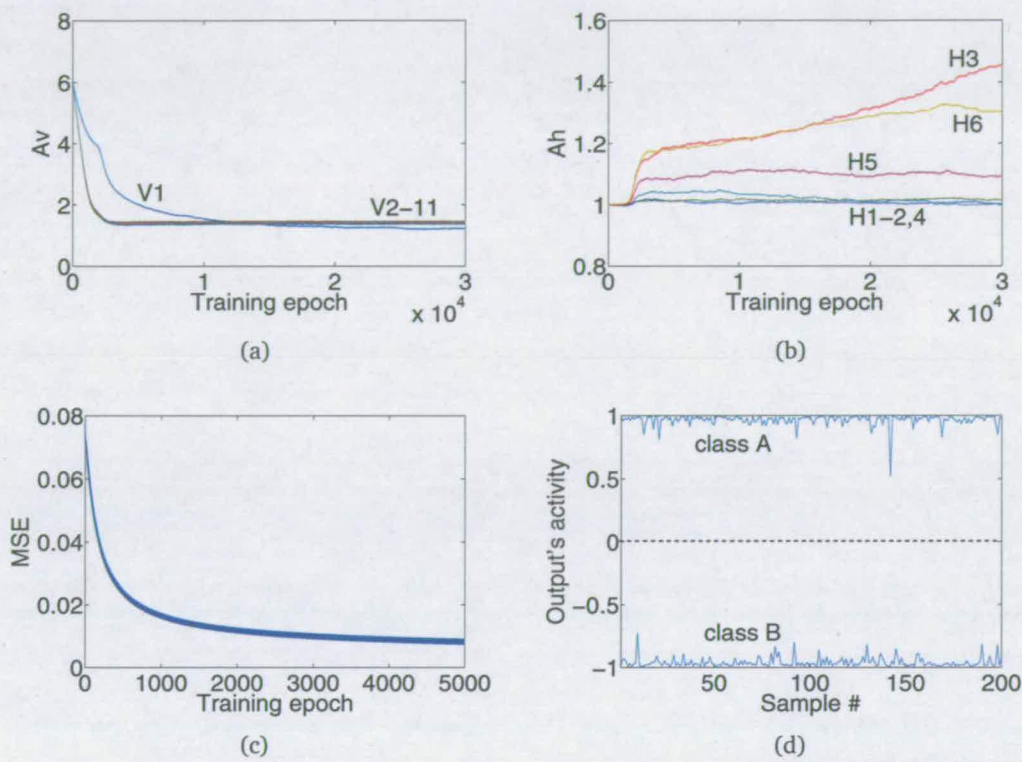
Figure 5.4: *Simulation results on multi-directional clusters: noise control parameter of visible layer A_v (a), noise control parameter of hidden layer A_h (b), MSE during the SLP training stage (c) and the SLP activity to 400 test samples (d).*

The initial high value ($= 6$) A_v was autonomously annealed to less than 1.46. In contrast, the noise control parameters A_h for hidden units H3 and H6 were increased to relatively more significant values 1.4519 and 1.2986, making them more sensitive to the class of input data (Fig.5.4(b) and Table 5.1). The dichotomy was further augmented via the SLP weight w_k (Table 5.1). The learnt model was tested with a new set of data, 200 samples for each class. Fig.5.4(d) shows the SLP activity with respect to the test data. Thresholding at zero, the neural system achieved 100% accuracy in data classification.

Parameter	Hidden unit					
	H1	H2	H3	H4	H5	H6
Final A_h	1.0024	1.0134	1.4519	0.9995	1.0923	1.2986
Final w_k	1.2154	-0.7434	7.9178	1.6897	3.3714	6.8491

Table 5.1: *Simulation results on the final noise control parameter A_h and SLP weight w_k.*
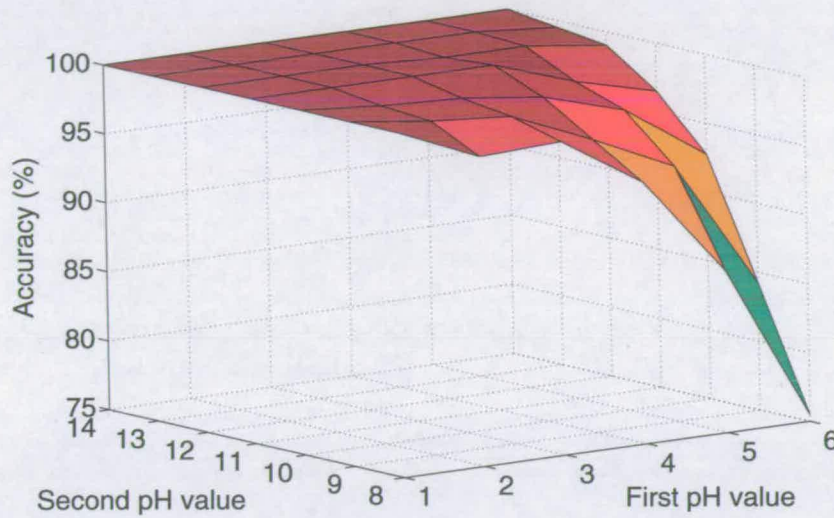
Figure 5.5: *Simulation results on different pairs (in pH values) of data clusters in terms of classification accuracy.*

Additionally, a series of simulations were run to examine the accuracy of the neural system on sensing different pairs (in pH values) of training data distributions. The temperature was fixed at a common point (37°C) for both clusters. The neural systems trained with each pair of data distributions are the same in terms of architecture size, noise scaling constants and learning rates. The test result was plotted onto Fig.5.5. As expected, the closer the two training data clusters were, the lower accuracy it achieved. The accuracy was 75.5% in the worst case where class A = pH 6 and class B = pH 8.

5.2 Classification trial two: meshed clusters

5.2.1 Modelling single non-Gaussian data distribution

This section explores if the neural system can classify 2D meshed data clusters, possibly non-Gaussian types. Therefore, it is important to establish and evaluate whether the CRBM can model single non-Gaussian data distribution. One example of such distribution with 392 data samples is shown in Fig.5.6.

A CRBM with 5 hidden neurons (Gaussian experts) was trained using 1-step Gibbs sampling methodology (Section 4.2) for 5000 epochs. The noise scaling constants of
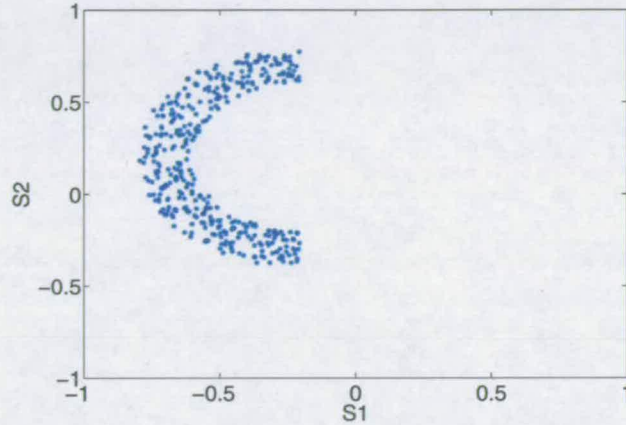
Figure 5.6: *A non-Gaussian, training data cluster.*

both the visible and hidden layers were set to 0.3. The learning rate of all parameters was 0.2. Fig.5.7 depicts the evolution of the noise scaling parameter A_v, with final A_v = {1.5000 2.8746}, whilst Fig.5.8 depicts the weight learning. Both figures suggested the CRBM has converged to a minima of its cost function.

However, the reconstruction of the CRBM with the learnt parameters was poor, as illustrated in Fig.5.9(a), indicating that the CRBM has only reached a local minima. In this simulation, there are three features to be learnt, namely the curve, the mean and the dispersion of the training distribution. As shown, the mean of the reconstruction was consistent to that of the training data distribution but the curvy feature has not been learnt. Obviously, this was due to poor learning in the noise component, $N_{vis} = A_v \times \sigma \times N_i(0,1)$, in the visible layer. There were two tunable parameters to refine the learning, namely the learning rate for A_v and σ. Both could be reduced to prevent the CRBM from being trapped at local minima, but at expense of longer learning time.

Alternately, the learning might be improved with a prolonged Gibbs sampling in each learning epoch to allow a better convergence because $KL(Q^1, Q^\infty) \geq KL(Q^n, Q^\infty)$ for $i \geq 1$. Thus, the previous simulation was repeated with a 20-steps Gibbs sampling in each learning epoch. In spite of the extra computational time, the reconstruction was equally poor (refer to Fig.5.9(b)). In a separate attempt, A_v was fixed at 0.1 throughout the 5000 training epochs. This time, a much improved reconstruction was obtained
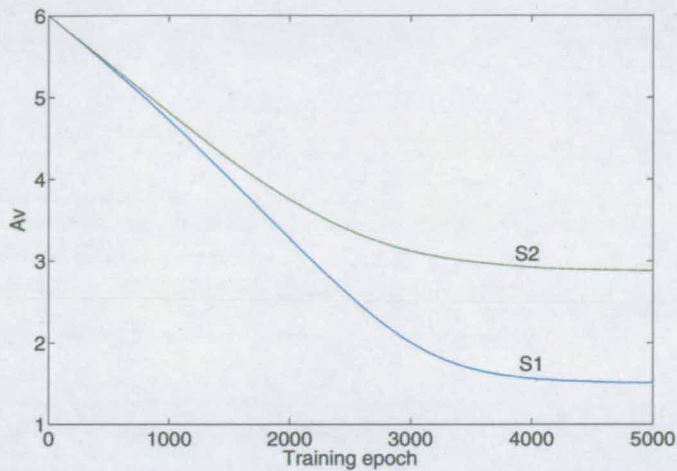
Figure 5.7: *The evolution of the noise scaling parameter A_v.*
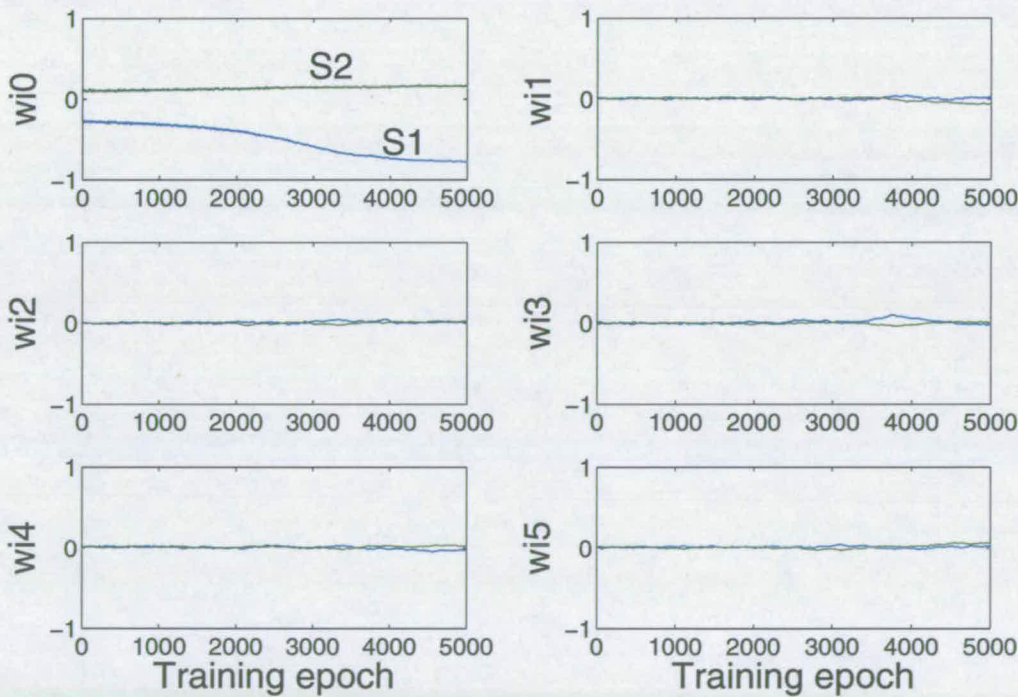
Figure 5.8: *The evolution of each hidden unit weight where w_{ij} refers to the weight between the visible neuron i and the hidden neuron j.*
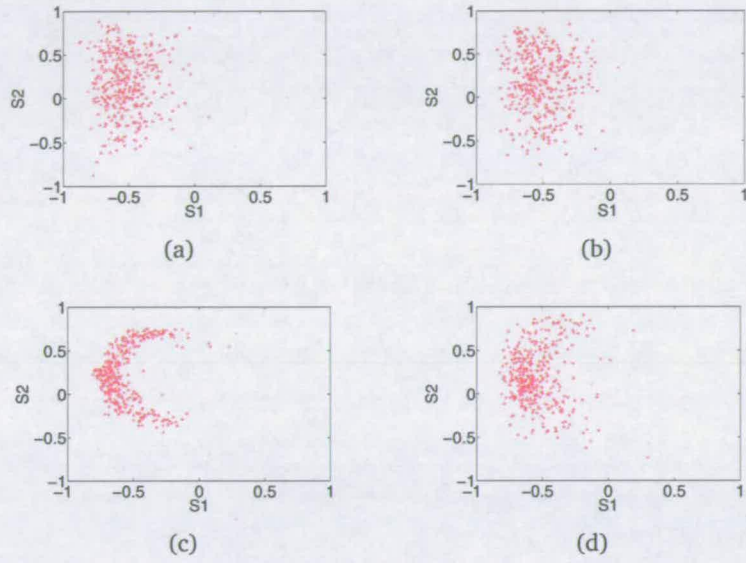
Figure 5.9: *Reconstruction for the CRBM with (a) learning A_v and 1-step Gibbs sampling, (b) learning A_v and 20-step Gibbs sampling, (c) $A_v = 0.1$ and 1-step Gibbs sampling, (d) $A_v = 1.0$ and 1-step Gibbs sampling, after 5000 training epochs.*

(Fig.5.9(c)). The reason for the improved result was that the noise component N_{vis}, which determined the dispersion of the distribution, has been scaled forcefully in this case. If A_v was much larger such as 1.0, the reconstruction would have been poorer (Fig.5.9(d)).

What has become apparent was $A_v = 0.1$ being in close proximity to the global minima of the cost function. To prove it, the previous experiment was repeated with (a) A_v initialised and fixed at 0.1 for the first 5000 training epochs then (b) A_v was free-running and its learning rate η_v was set to 0.01 for the next 5000 training epochs. Such small η_v could prevent any unwanted weight saturation due to over-large learning steps. Fig.5.10 depicts the simulation result on A_v over the entire training period. Despite the initial spike around 5000th training epoch, A_v settled eventually at approximately 0.115 for both visible units S1 and S2. A smaller η_v could reduce the amplitude of the oscillation but at the price of much longer training epochs (> 100000). No significant change was observed in the weights and A_h of the CRBM in the second 5000 training epochs, hence the reconstruction at 10000th training epoch was similar to that of 5000th's.

From the above simulations, it could be concluded that it is possible for the CRBM to
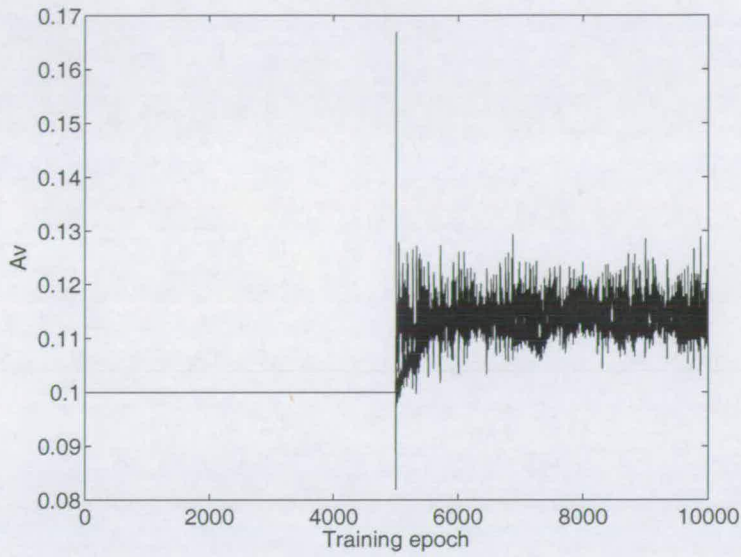
Figure 5.10: *Parameter A_v in time series for a CRBM with first fixed then learning A_v and 1-step Gibbs sampling.*

encode a non-Gaussian data distribution reasonably well (Fig.5.9(c)), despite all of the experts being Gaussian. More importantly, the CRBM must have very small the learning rate for A_v to avoid trapping at a local minima, and be trained over a long period of time to permit it to converge to global minima of its cost function. To speed up the simulation process, A_v was fixed for subsequent simulations.

5.2.2 Simulation one

Two simulations were performed to examine if the neural system could classify 2D meshed data clusters. In the first simulation, a CRBM with 7 hidden neurons was trained with a 1-step Gibbs sampling for 5000 epochs to encode the training data distributions with 400 samples for each cluster, as illustrated in Fig.5.11(a). The noise scaling constants were fixed at 0.4. Noise control parameter A_v at the visible layer was held constant at 0.1 while noise control parameter A_h at the hidden layer was allowed to evolve into feature detectors. All the learning rates were set to 0.2.

The training data distributions were considerably complex to be modelled by the CRBM, especially with common noise control parameters and noise scaling constants, because the two clusters had different dispersions and shapes. Fig.5.11 depicts the reconstruc-

tion data distributions after the learning stage, which exhibited some resemblance to the training data distributions. To quantify the performance, a SLP was trained for 5000 epochs to linearly classify the response in the CRBM, more precisely the activity of the hidden units, with respect to the class of input data: class A data yielded an output of "+1"; class B data gave an output of "-1". Fig.5.11(c) shows the SLP response to a set of test data, 400 samples for each class. Thresholding the response at zero, the trained neural system showed a hundred percent of accuracy.

5.2.3   Simulation two

To further validate the capability of the neural system, a second set of meshed clusters were employed for the similar simulation. This time both training data clusters were non-Gaussian (refer to Fig.5.12). Due to the increase in complexity, the CRBM was trained for longer time, 20000 learning epochs to be exact. Upon completion, a 20-step Gibbs sampling reconstruction was generated as shown in Fig.5.12(b). Repeated as before, a SLP was used to evaluate the performance of the non-linear data modelling by the CRBM. A high accuracy (99.7416 %) in test data classification was achieved. The result has demonstrated once again the strength of the CRBM in data modelling despite of its relatively simple and small architecture.

5.2.4   Performance evaluation: number of training epoch

Thus far, the CRBM has demonstrated its ability to model two meshed, non-Gaussian data distributions simultaneously. In this and the next section, temporal and spatial

(a)   (b)   (c)
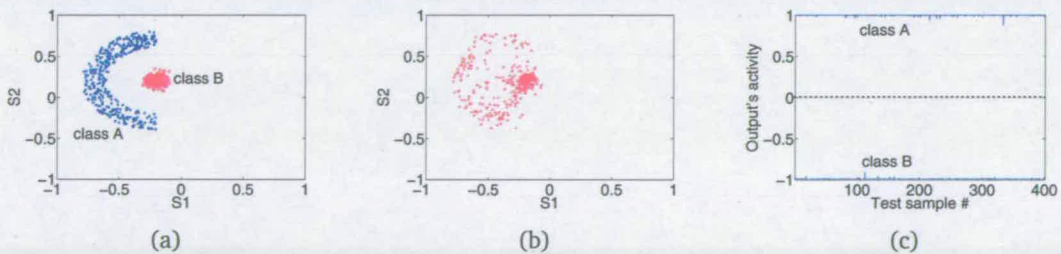
Figure 5.11: *Simulation one on meshed clusters: (a) training data, (b) reconstruction data after 5000 epochs and (c) the SLP activity with respect to a set of test data.*
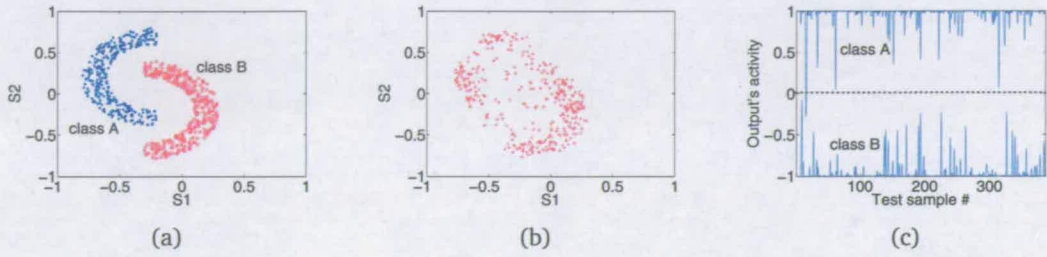
Figure 5.12: *Simulation two on meshed clusters: (a) training data, (b) reconstruction data after 20000 epochs and (c) the SLP activity with respect to a set of test data.*

ways to optimize the CRBM to model the training data distribution (Fig.5.12(a)) are investigated. In this context, it referred to performance evaluation on data classification with different numbers of training epochs and with different numbers of hidden units, using common noise scaling constants and learning rates.

A CRBM with 7 hidden units was trained over fourteen different numbers of epochs {100, 200, 300, 400, 500, 750, 1000, 2000, 5000, 10000, 20000, 30000, 40000, 50000}. In all the simulations, the learning rate for the noise control parameter at the hidden layer was fixed at 0.2, and the noise scaling factor was 0.4 for both visible and hidden layers. The noise control parameter at the visible layer was fixed at 0.1 to reduce the simulation time. As previous, the modelling performance of the CRBM was evaluated by a linear classifier which would be trained for 5000 epochs at a rate of 0.2.

The set of simulations was run three times, each with a different learning rate (η = {0.01, 0.05, 0.20}) for the CRBM weight. The simulation results were plotted in Fig.5.13. When $\eta = 0.20$, the convergence was at the fastest rate, achieving approximately 100% accuracy within 500 training epochs. However, it did not settle down, it oscillated over time due to the large learning step size. In comparison, when $\eta = 0.05$, the convergence was more stable in spite of the slower rate (achieved 100% accuracy only after 1000 training epochs). The slowest, but also the smoothest convergence, was achieved at $\eta = 0.01$ as expected. In each learning epoch, the model moved in a very small step towards the training data distributions. After 10000 epochs, the model eventually achieved 100% accuracy in test data classification.
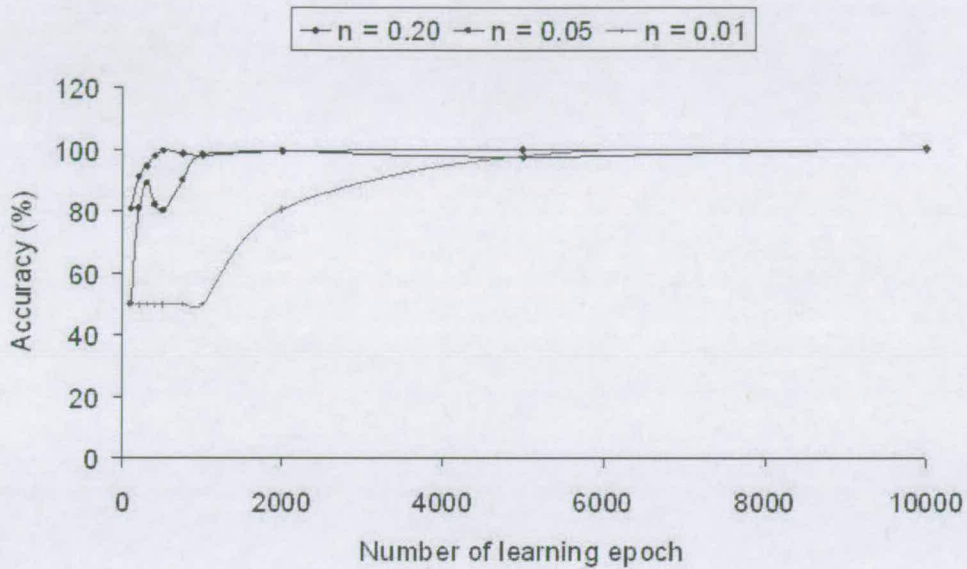
Figure 5.13: *The neural system accuracy on data classification with various numbers of training epoch. The term 'n' is referred to the learning rate of the CRBM, in this diagram.*

5.2.5 Performance evaluation: number of hidden units

Generally, the number of hidden units required to model the training distributions depended on the complexity. The more features the distributions possessed, the more hidden units would be necessary. Previous sections assumed that the CRBM could capture the key characteristics of the training distributions with 7 hidden units. However, the aim, in this section, is to determine the minimum number of hidden units which are sufficient to fulfill the application requirement - mapping non-linear, meshed data vectors into linearly-separable spaces.

A CRBM with different numbers of hidden units ($= \{1, 2, 3, ..., 16\}$) was trained to model the data distributions (refer to Fig.5.12(a)) for 20000 epochs. The learning rates were all set at 0.2, while the noise control parameter of the visible layer was fixed at 0.1 and the noise scaling factors were 0.4. For each model size (or number of hidden units), a SLP was trained at a rate of 0.2 for 5000 epochs to provide a quantitative evaluation on how well the CRBM has learnt. Subsequently, the trained model was tested with a new set of data, 200 samples for each class. These training and testing exercises were run three times to affirm the consistency of the simulation results.

The simulation results for each model size were captured and plotted in Fig.5.14. As expected, the single hidden neuron faced some difficulties to project such complex distributions into a linearly separable space. As a result, a poor data classification - merely 76.3% accurate was produced. The modelling power was immediately improved with the introduction of the second hidden unit, which threby improved the accuracy to 96.3%. No significant further improvement was observed, on average, with the third addition of hidden unit. Eventually, when the model size was increased to 5 hidden units, approximately 100% accuracy was achieved. Further incremental increases in the model size has resulted insignificant variation in the accuracy ($\pm0.173\%$) on test data classification. The optimum model size for the pair of training distributions was thus five hidden neurons.
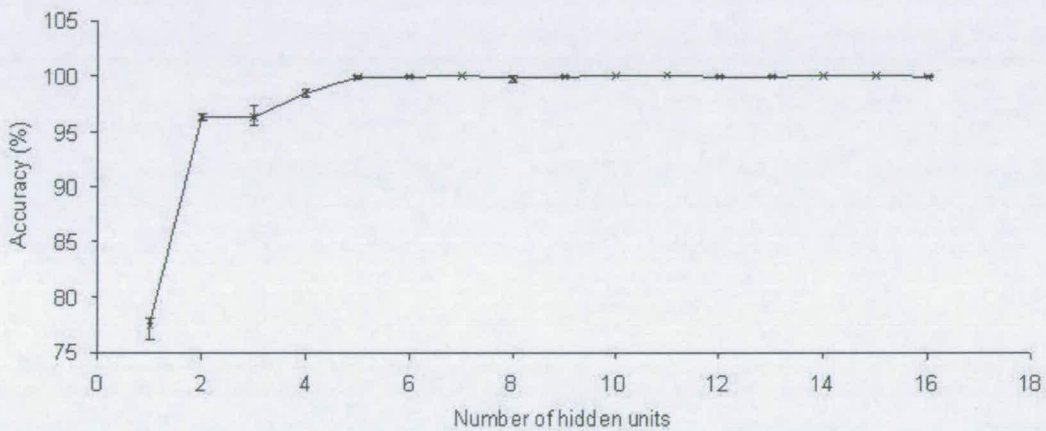
Figure 5.14: *The neural system accuracy on data classification with various numbers of hidden neuron.*

5.3 Discussion

Overlapping clusters are common in pattern recognition applications such as odour detection and electroencephalogram (EEG) classification. The overlap is due to limited sensory sensitivity and noisy environment. In [43], Shin tried three different neural algorithms, namely MLP, Learning Vector Quantization (LVQ) and Fuzzy ARTMAP, to preprocess the transient response of gas sensors before passing the more linearly-separable features as inputs to a PCA to classify. Alternatively, Sarry [41] employed Discriminant Factoral Analysis (DFA) to extract the features and then PCA to classify. Other pre-

processing techniques employed in the literature are self-organizing maps (SOM) [37], Genetic Algorithm (GA) [40], Wavelet Transform [148] and RBF [149]. While most of them provided very promising results, they are in general not suitable for the *IDEAS* project in terms of hardware-amenability as discussed in Chapter 2.

In Section 5.1, few simulations have been run with CRBM extracting the main features in 2-D and multi-dimensional overlapping data clusters. The data distributions took a Gaussian form and were based on the sensor models derived in Chapter 4. The simulation results demonstrated that the CRBM could extract the features required to allow the subsequent-stage SLP to distinguish the two clusters effectively, achieving 100% accuracy with test datasets. It has also observed that when more redundant sensors were employed, more information could be fused, giving more chances to the neural system to detect the dichotomy. As a result, the SLP response to high dimensional data clusters was more distinguishable. In short, a higher dimensional sensory system is more robust, similar to our own biological systems such as tongue and skin.

The simulations were then extended to classification problems with non-Gaussian, meshed data distributions. The way CRBM encoded a single non-Gaussian data distribution was first examined. The noise scaling parameter in the visible layer, A_v, was identified as the key player in the data modelling. When the complexity of the data distribution is increased, A_v must be small in value to ensure that the trained CRBM has encoded the segmentation between the clusters and could reconstruct each cluster in high precision. However, this means that a longer training period is expected before a convergence to the global minima of the cost function can be achieved.

Subsequently, two simulations were run (refer to Section 5.2.2 and 5.2.3), both on non-Gaussian, meshed data classification problems. The simulation results demonstrated that the CRBM could model such distributions and classify the data with high accuracy. As expected, the difficulty in attaining 100% accuracy escalates with the complexity of the data distributions; evidential in the pair of simulation results. The response of the neural system to the two clusters was more distinguishable in the first simulation than the second.

The above simulation results deduced that the neural system could be used as a robust classifier and it is applicable even for non-Gaussian, meshed data classification

problems. The limitations of the model have also been investigated with reference to several issues as follows.

Overlapping/meshed clusters The accuracy in classification of the neural system is primarily confined by the training data distributions. When the two clusters are overlapping each other or have meshed regions, the performance of the model will naturally degrade just like any other computational models. However, as in Section 5.1.2, its robustness could be improved by introducing more redundant sensors.

Convergence rate vs stability Section 5.2.4 searched for the number of training epochs necessary for the neural system to achieve its aim - 100% accuracy in data classification. It was observed that large learning rate η allowed faster convergence but less stability - i.e. the trained model could not consistently produce full accuracy. In contrast, small η reached the optimum solution much later, possibly 3-4 times longer training period, yet the trained model was more stable in reproducing full accuracy results. Therefore, it is the trade-off between the convergence rate and the stability of the final model. Additionally, it is important to select the right η for complex (meshed) data distributions, in particular to keep the model away from any local minima during its evolution. More detailed discussion about the selection of η could be found in [150].

Number of hidden units The simulation results in Section 5.2.5 showed that a "sufficient" number of hidden units must be supplied to secure high accuracy in data classification. However, it was not possible to know the optimal number of hidden units without the process of trial and error. The sufficiency depends on the number of features embedded in the data distributions. The higher the number, the more hidden units are required to encode them. When there is a lack of hidden units, the neural system cannot classify data accurately and consistently. On the other hand, when there are too many hidden units, those extra units will either compete with the main units[1] for a share of the main features - i.e. a feature may be jointly coded by two or more hidden units, or simply be idle, storing no information. Although high accuracy in data classification is likely to be guaranteed,

---

[1] In this context, the main units refer to those hidden units that encode the main features of the training data distributions.

having extra hidden neurons is undesirable in terms of hardware implementation because they occupy extra silicon area and consume power.

5.4 Summary

In the first part of this chapter, the mechanism of the neural system (refer to Chapter 4) to classify two clusters of data was explained. The first example distributions were simple but overlapping in two dimensions, while the second example distributions had their dimentionality increased to eleven. Both examples had their data derived from the sensor models in Chapter 3, and had Gaussian distributions. Full accuracy in data classification for the trained model was recorded.. Simulations with different sets of environmental conditions were performed in Section 5.1.2. While the two clusters (pH 6 and 8) were heavily overlapping each other, the trained model still managed to achieve an accuracy of 75.5%.

In the second part of this chapter, the neural system was applied to non-Gaussian, meshed data classification problems. Two different simulations were run and their results illustrated the strong modelling capability of the model. Besides that, the convergence rate and the stability of the model have been investigated in Section 5.2.4, and the optimal number of hidden units needed to encode all the features in these particular data distributions was identified (Section 5.2.5). A discussion on the overall simulation results concluded this chapter.

Chapter 6

Classification simulation on real drifting data

In this chapter, the neural system is applied in a dynamic environment, where stochastic sensory drift degrades the data classification. With no control measurements is available during the drift experiment, supervised neural algorithms are normally unsuitable for this application. However, complex compensation scheme can be introduced. An example is demonstrated with a MLP in [151]. Another major challenge in this drift experiment is the absence of data from one of the two classes over a lengthy measurement time. Neural algorithms such as Kohonen's SOM [35, 37] will face major difficulties to balance between learning the drift and avoiding CI. The main objective of this chapter is to demonstrate that a simple neural algorithm such as the forementioned neural system (Chapter 4) can auto-calibrate against such drift even in high-dimensional, meshed clusters problems without experiencing CI. To do so, the neural system is tested in two separate drift trials, one targeted for high-dimensional problems whilst the other for meshed clusters problems. Section 6.1 will present the experiment to record real drifting data which will be used in subsequent simulations. Sections 6.2 and 6.3 will describe the two separate drift trials. From the simulation results, the limitations of the neural system will be identified and discussed.

6.1 Preparation of drifting dataset

6.1.1 Experiment setup

Long term pH measurements were carried out by Erik Johannessen to assess signal drift and sensor lifetime in a physiological electrolyte (0.9% saline) solution of pH 7. Its ambient temperature was maintained at 36.5°C throughout the experiment by using a water bath and having the assay solution continuously stirred and re-circulated by a peristaltic pump. Its pH value was closely monitored by a reference pH meter (Consort
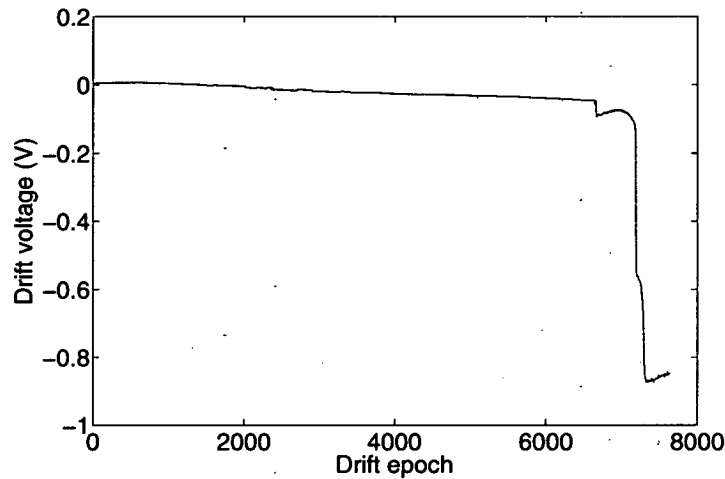
Figure 6.1: *Drift data in time series. Note that the pH-ISFET sensor has a sensitivity of -23mV/pH. Thus, for instance, a drift voltage of -163.8mV will result a drift of +7 pH units.*

n.v., Belgium). The signal conditioning circuit for the pH-ISFET was as in Fig.3.5(b) and was powered by $\pm1.55V$ sources. The Ag|AgCl reference electrode, representing the potential in which the floating gate was referred to, was connected to ground. The total duration of the experiment was 21h.

6.1.2 Example data

Throughout the experiment, the reference pH meter reported a constant pH value for the solution. Therefore, any drift recorded at the output of the signal conditioner would be due to the instability of the Ag|AgCl reference electrode. Fig.6.1 illustrates a set of measurements in time series with pH 7 calibrated at 0V. Each drift epoch corresponds to 10s of sampling period.

A gentle but stochastic drift of $-8\mu V$/epoch ($R^2 = 0.9833$) was recorded during the first 6600 drift epochs, in which a total drop of -0.047V occurred. The drift was primarily due to the dissolution of AgCl from the reference electrode. Then, there was a sudden drop of voltage. The seal of Nafion$^{\circledR}$ membrane was broken and the reference potassium chloride gel electrolyte solution which was previously encapsulated, then leaked into the measuring environment, causing a failure in the reference electrode.

When the experiment was repeated on other reference electrodes, the drift patterns were similar in general but the seals of the Nafion membrane were broken at different time instants due to packaging variation.

6.2 Drift trial one: High dimensional, simple clusters

In this trial, the neural system in Chapter 4 was first trained with some static data. The trained system was then fed with drifting (dynamic) data, one sample per drift epoch. The system was tested with a group of samples at four time intervals, namely the start, the 4000th epoch, the 6000th epoch and the end of the drift simulation. The weight changes in the CRBM were also recorded throughout the experiment for examination if the model could auto-compensate sensor drift.

6.2.1 Simulation dataset

In this case, the neural system was applied as an adaptive stochastic classifier in a multisensor microsystem. It had one temperature (S1) and ten pH-ISFET (S2-11) sensors. Four reference electrodes (R1-4) were used to improve the robustness of the overall measurement system and were allocated arbitrarily to different sensors, i.e. R1 for S2-8, R2 for S9, R3 for S10 and R4 for S11. Fig.6.2 depicts the architecture of the forementioned microsystem.

The static dataset used to train the model was similar to that in Section 5.1.2 (refer to Fig.5.3(a)-(c)). The two clusters were ovelapping each other and consisted of 200 samples each. Due to its high dimensionality, they were linearly separable as seen later in Section 6.2.2. On the other hand, the dataset used under a dynamic environment was based on the drift pattern in Fig.6.3. Each reference electrode displayed different drift and failure rates. There were three interesting time intervals over the 7644 drift epochs. Before the 4000th epoch, all sensors except S9 were experiencing minor drift. Between the 4000th and the 6000th epochs, the reference electrode R3 broke down and hence S10 failed. After the 6000th epoch, all remaining reference electrodes were degrading until all the sensors become faulty eventually. Over the whole drift period, the temperature sensor S1 was assumed to be functioning correctly and in stable. Only samples from *Class A* dataset were available for the model to autocalibrate itself.
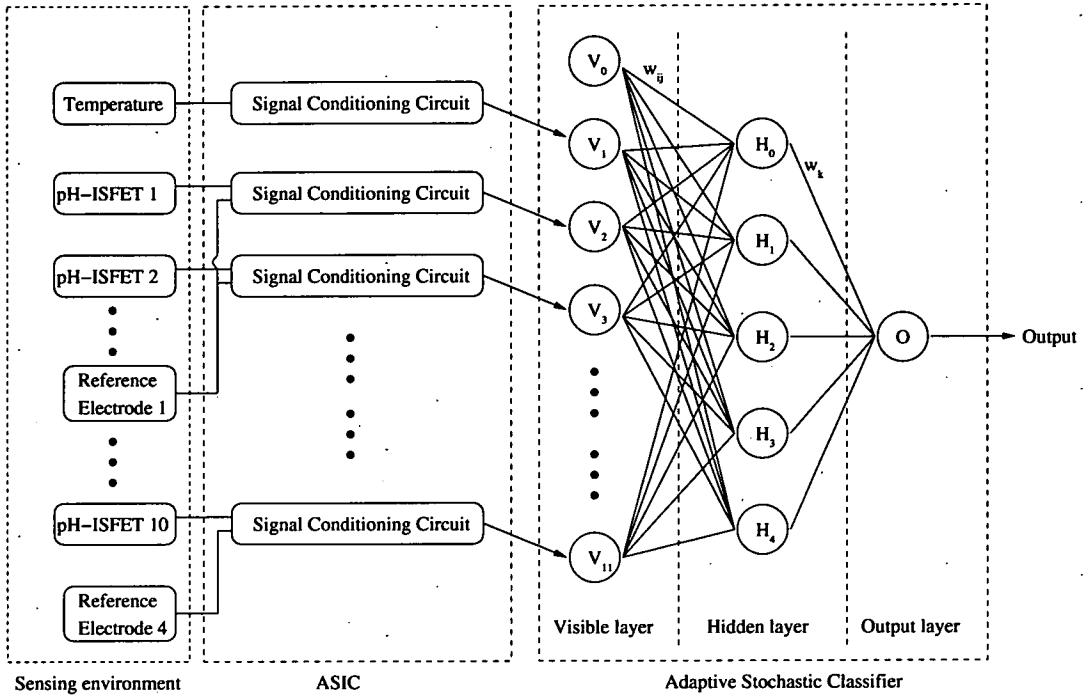
Figure 6.2: *System architecture of an 11-dimensional microsystem.*
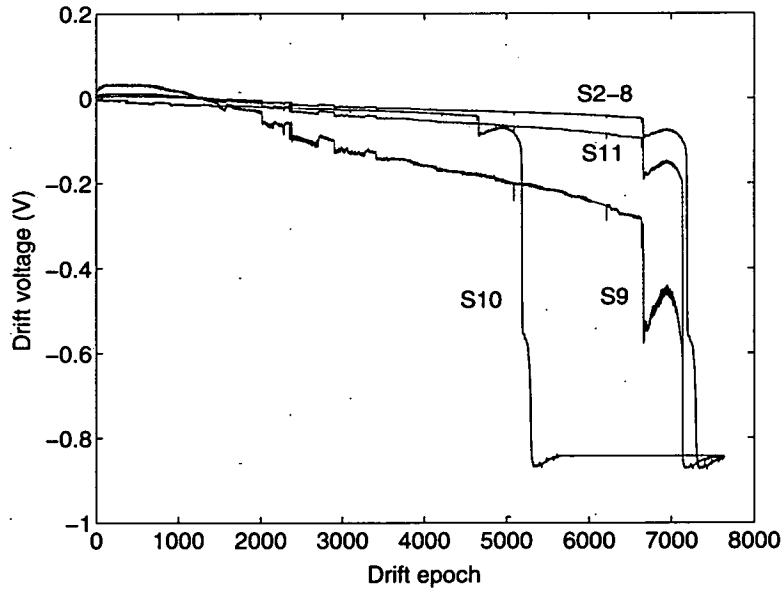
Figure 6.3: *Drift voltages for the ten pH-ISFETs (S2-11).*

6.2.2 Simulation result

Initially, the neural system was trained at 20000 epochs for the CRBM and 5000 epochs for the SLP. A MSE of 7.400×10^{-3} was achieved upon the completion of the training, as illustrated by Fig.6.4. More discriminative weights of the SLP, i.e. smaller MSE, could be achieved by extending the training period. Despite that, a classification accuracy of 100% was attainable at this stage when the trained system was tested with 200 new samples from each class (see Fig.6.6(a)).

Next, the trained system was fed with the drifting data (Fig.6.3), one drifting sample per epoch. Throughout the 7644 drift epochs, the CRBM in the neural system was allowed to adapt to the sensor drift as described in Section 4.2.3. The weights of the bias units in both visible w_{0j} and hidden layers w_{i0} were updated at a rate η_w of 0.01. Fig.6.5 depicts the weight changes in w_{i0} for pH-ISFET sensors Si where $i = \{2, 3, ..., 11\}$. The changes correlated to the drift patterns in Fig.6.3, indicating that the model could track autonomously the drift. Based on the updated w_{i0}, corresponding weight changes in w_{0j} were implemented to ensure the activation of the hidden units (especially the dichotomous detector) were thresholded correctly. Evidently, the model managed to score 100% in data classification accuracy when tested at 4000th and 6000th drift epochs (Fig.6.6). However, the performance of the model deteriorated drastically near the end of the drift simulation, suggesting that the model has broken down completely and was then, quite rightly, responding to the input signal as part of a completely new data distribution. The drastic performance degradation is caused by the simultaneous failures in all the remaining pH-ISFET sensors.

To highlight the significance of the online unsupervised learning in the neural system, trained but subsequently non-adaptive linear (another SLP) and MLP classifiers were used as benchmarks. Both were initially trained with the same datasets (without drift) as before, but their weights were not updated online during the entire drift period.

The SLP had a sigmoidal activation function and was batch trained with the Widrow-Hoff delta rule[144] for 20000 epochs. The learning rate for its weight was 0.05. Fig.6.7(a) illustrates the evolution of the SLP learning in terms of MSE. The learning was completed with a final MSE 7.895×10^{-3}. Then the trained SLP was fed with the drifting data. When tested with 200 samples from each class, the SLP showed 100%
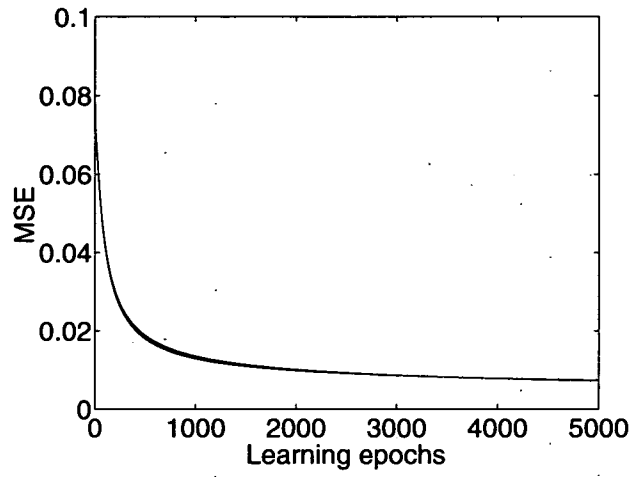
Figure 6.4: *The evolution of the SLP learning in the neural system in terms of MSE.*
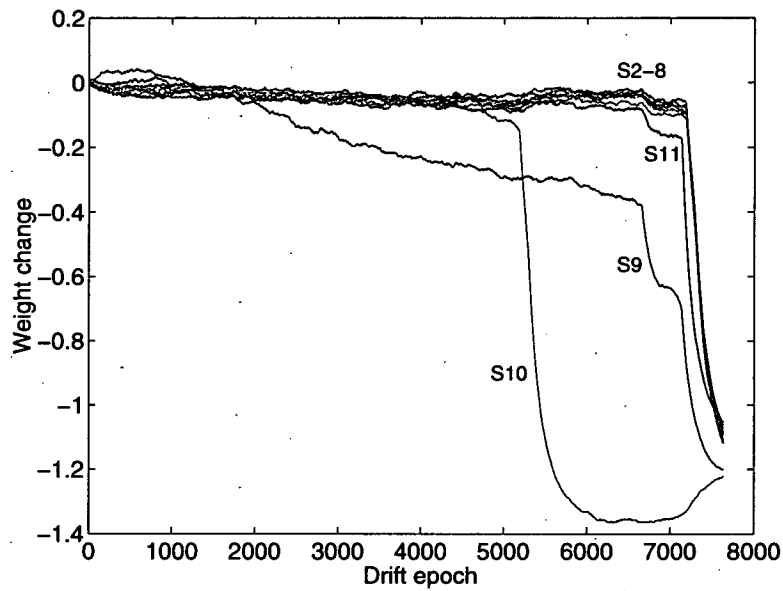
Figure 6.5: *The weight changes in the hidden bias unit in response to the sensory drifts in the ten pH-ISFETs.*
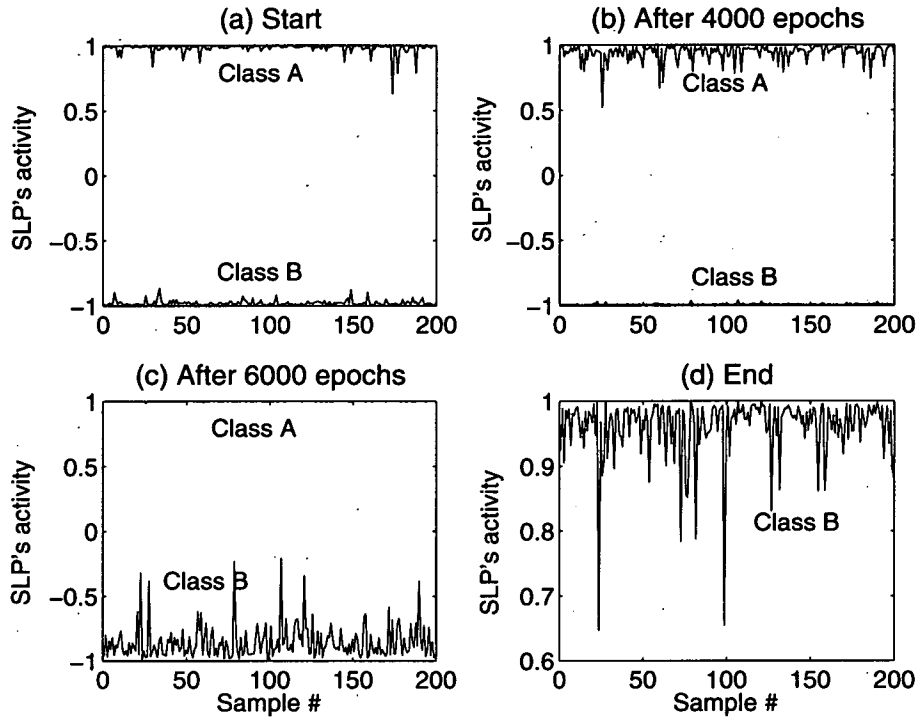
Figure 6.6: *The classification accuracy of the neural system at different drift epochs.*

accuracy in data classification at the 4000th drift epoch. However, the performance deteriorated dramastically in subsequent epochs, primarily due to the sudden failure of sensor $S10$ and the relatively large drift experienced by sensor $S9$. The accuracy dropped to 53% when tested at the 6000th drift epoch. As seen in Fig.6.8(c), the SLP was biased towards class A. Such a result was expected because the two clusters (refer to Fig.5.3(c)) had drifted towards the negative quadrants for both axes, according to the drift patterns in Fig.6.3. Eventually, both clusters would be on the "original class A side" of the hyperplane defined by the weights of the SLP, thus resulting the 50% classification accuracy.

Similar results were encountered by the MLP. It had an architecture of 7:7:1 and each neuron had a sigmoidal activation function. It was initially batch trained with a "back-propagation" learning rule (based on gradient descent with momentum [152]). The total learning duration was 20000 epochs, the learning rate was 0.05 and the momentum was 0.9. As foreseen, the MLP achieved a better learning with a final MSE 1.446×10^{-4} (refer to Fig.6.7(b)) than the SLP. Further evidence is depicted in Fig.6.9(a), when the

(a) SLP

(b) MLP
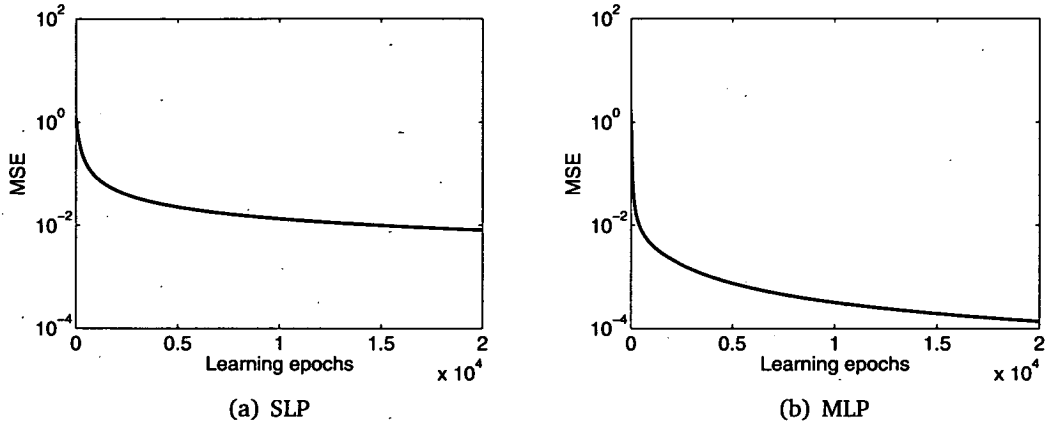
Figure 6.7: *The evolution of SLP and MLP learning.*
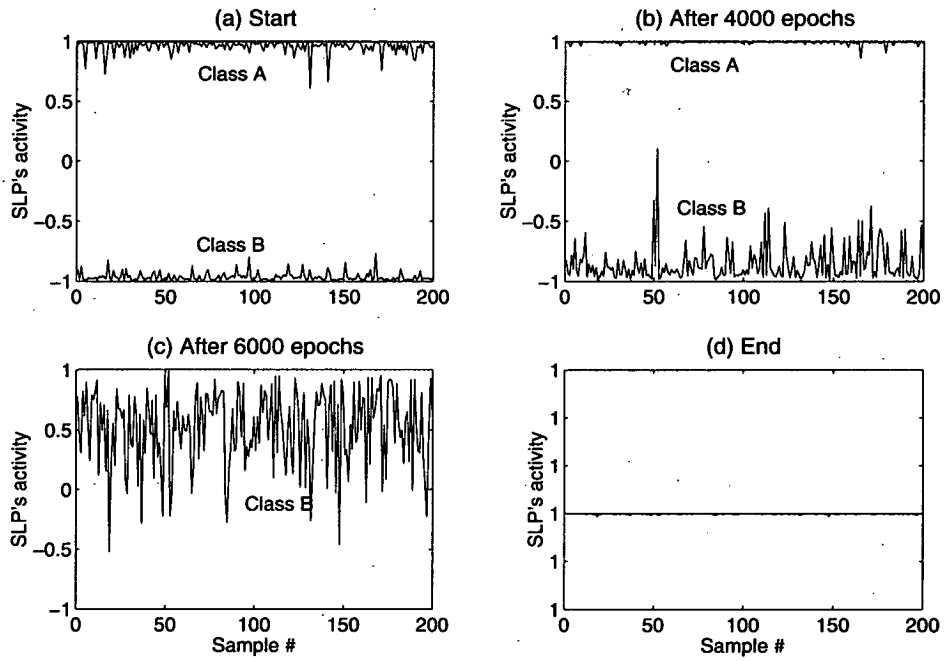
Figure 6.8: *The classification accuracy of the SLP at different drift epochs.*
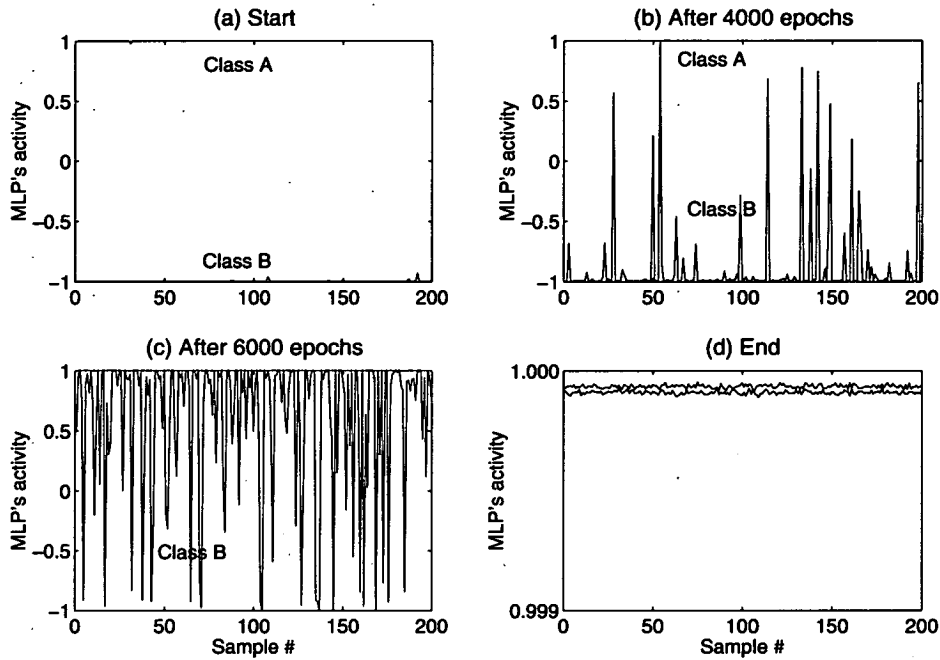
Figure 6.9: *The classification accuracy of the MLP at different drift epochs.*

trained MLP tested with 200 samples from each class. Subsequently, the trained MLP was fed with drifting data (Fig.6.3) without any further weight updating (i.e. drift was not compensated). For the first 4000 drift epochs, the classification accuracy was still 100% but there was a sign of degradation in performance. The MLP response to class B test data was less saturated to "-1" in Fig.6.9(b) than a. When tested again at the 6000th drift epoch, the accuracy was down to 58%. Just like the SLP, the MLP was biased towards class A but quicker. This is probably because the discriminating hyperplane is better defined with its weights. The summary on the classification accuracy of the above three algorithms at different drift epochs is tabulated in Table 6.1.

Algorithm	Accuracy (%) at drift epoch #			
	Start	4000th	6000th	End
CRBM+SLP	100.00	100.00	100.00	50.00
SLP	100.00	99.75	53.00	50.00
MLP	100.00	97.75	58.00	50.00

Table 6.1: *Simulation results on the classification accuracy at different drift epochs by different algorithms.*

6.3 Drift trial two: Two dimensional, meshed clusters

The previous trial has looked at a high-dimensional, simple yet overlapping clusters classification problem, in a dynamic environment. Similar experiments were repeated in this trial but with a pair of two-dimensional, meshed clusters. The objective was to demonstrate that the scheme – the neural system and its learning methodology – proposed in this thesis was also applicable to non-linear classification problems.

6.3.1 Simulation dataset

A 6-hidden-unit CRBM of the neural system was fed with two outputs from sensors S1 and S2. The training dataset in a static environment for these two dimensions was similar to that in Fig.5.12(a). The meshed area between the two clusters poised itself as a non-linear classification problem. The two sensors had a reference electrode each. In this trial, the reference electrodes experienced similar but opposite direction drifts, which shared a pattern as in Fig.6.1. The drifts moved the two clusters towards the positive quadrant in S1-axis and the negative quadrant in S2-axis such that cluster A data would be in the original position of cluster B after T number of drift epochs, where $T \gg 1$ and $T \propto 1/(driftrate)$. To accentuate the online learning by the CRBM, the drift rate used in this simulation was assumed to be six times greater than that in Fig.6.1. Similarly to before, only samples from Class A dataset were available throughout the entire drift period.

6.3.2 Simulation result

The CRBM was first trained, with the forementioned datasets in a static environment for 20000 epochs before the SLP for another 5000 epochs. A MSE of 4.800×10^{-3} was achieved upon the completion in learning. The trained neural system was able to classify in 100% accuracy 200 test samples from each class, as depicted in Fig.6.10(a). Subsequently, the model was fed with drifting data from Class A, one sample per epoch. The CRBM weights were allowed, in a constrained manner, to be updated at a rate of 0.01. After 4000 drift epochs, the updated model was reported to maintain an accuracy of 96.28% when it was tested with 200 samples from each class. As time went by the drift shifted the data distributions farther away from their original positions in S1/S2-
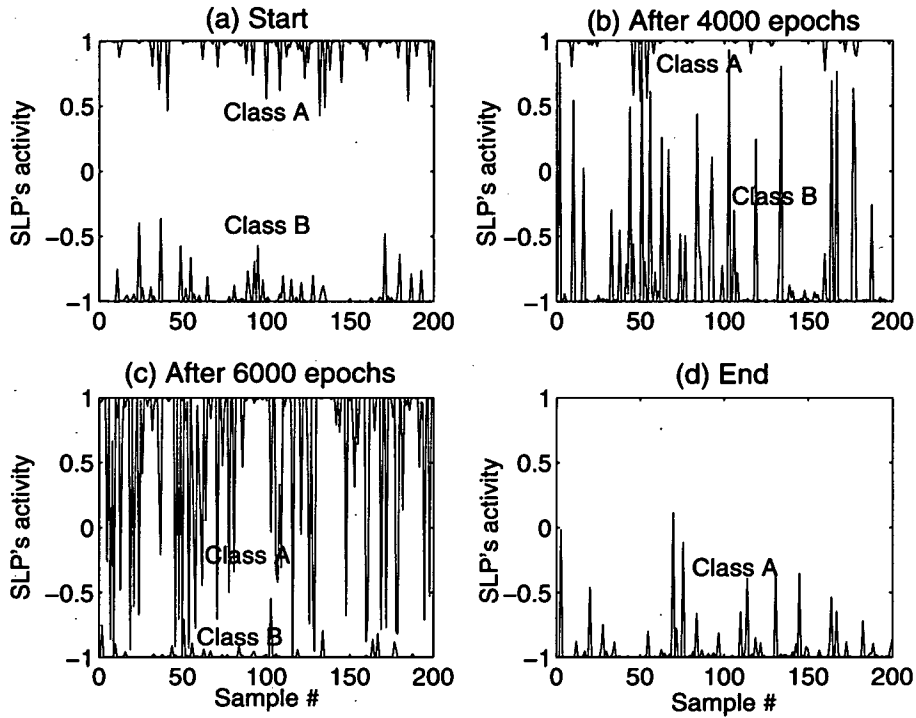
Figure 6.10: *The classification accuracy of the neural system at different drift epochs.*

axes. At the 6000th drift epoch, the model become more biased towards Class B (refer to Fig.6.10(c)), as anticipated. When tested with new samples, the model managed to classify with 89.83% accuracy. By the end of the drift period, both the reference electrodes failed, hence caused, as expected, a huge drop in classification accuracy (Fig.6.10(d)).

To highlight the adaptivity feature of the CRBM, the above simulation was repeated with another SLP and MLP. The SLP was trained in a static environment for 20000 epochs. With the Widrow-Hoff delta rule, the final MSE was 1.378×10^{-1}. Fig.6.11(a) illustrates the SLP activity/response to 200 test samples from each class. The classification accuracy was 95.20%, indicating the problem was a non-linear one. Later, the trained SLP was fed with the same drifting data as those for the neural system. No online learning was allowed during the course of 7644 drift epochs. When tested with samples, its classification accuracy was dropped to 86.53% after 4000 epochs. Refer to Fig.6.11(b), the SLP was obviously biased towards Class B. Further temporal drift exacerbated the classification accuracy. Evidently, it became 78.80% when the SLP was
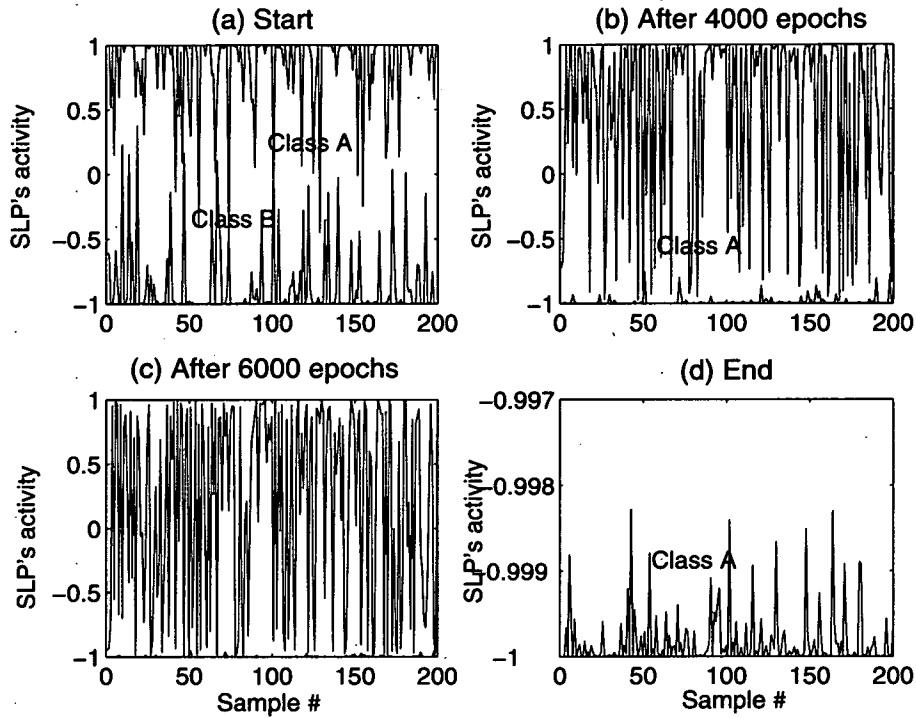
Figure 6.11: *The classification accuracy of the SLP at different drift epochs.*

tested after 6000 epochs (Fig.6.11(c)) before it collapsed eventually.

Next, the same simulation was re-run with a non-linear classifier - a MLP. It posssesed an architecture of 7:7:1 and each neuron had a sigmoidal activation function. It was trained with a backpropagation learning rule (gradient descent with momentum) for 20000 epochs. The final MSE was 1.102×10^{-4} and the trained MLP was able to classify test samples 100% accurate (refer to Fig.6.12(a)), hence outperforming the single SLP. The MLP was then examined with the drifting data. After 4000 epochs, a classification accuracy of 85.66% was recorded when the non-adaptive MLP was tested with drift samples. As in the SLP case, the MLP was found to be biasing towards Class B (Fig.6.12(b)). Its accuracy degraded further to 74.18% at the 6000th drift epoch. The bias becomes clear in Fig.6.12(d) which shows the classification result at the end of the drift period. The MLP output is always below -0.99, thus producing false labels on Class A drift data. A summary of the classification accuracy of the three algorithms at four different drift epochs is listed in Table 6.2.

Beside the examination with test samples at different drift epochs, the outputs of each
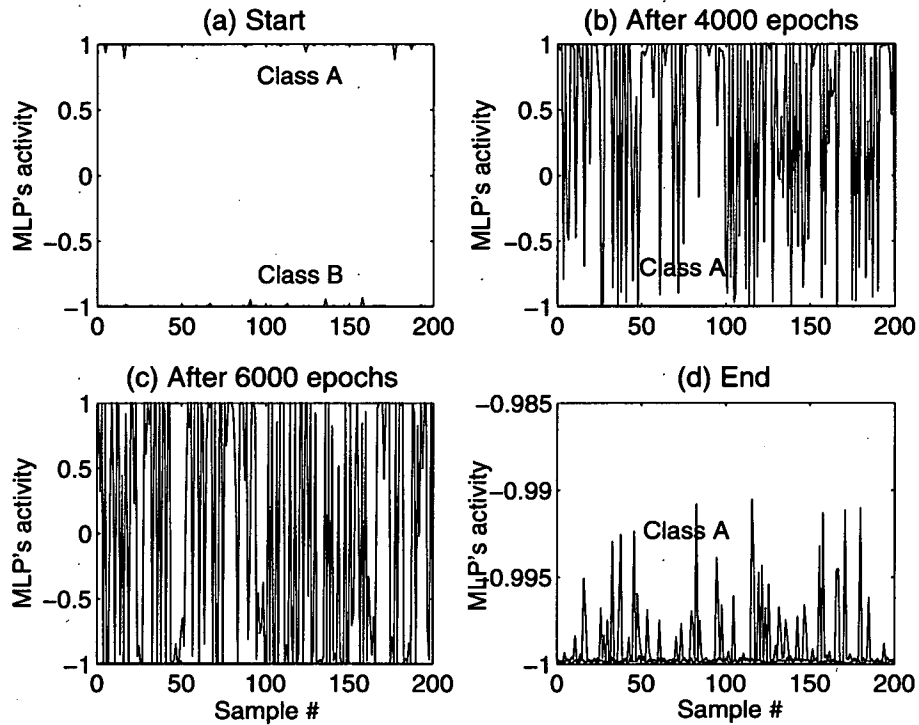
Figure 6.12: *The classification accuracy of the MLP at different drift epochs.*

	Accuracy (%) at drift epoch #			
Algorithm	Start	4000th	6000th	End
CRBM+SLP	100.00	96.28	89.83	50.37
SLP	95.20	86.53	78.80	50.00
MLP	100.00	85.66	74.18	50.00

Table 6.2: *Simulation results on the classification accuracy at different drift epochs by different algorithms.*
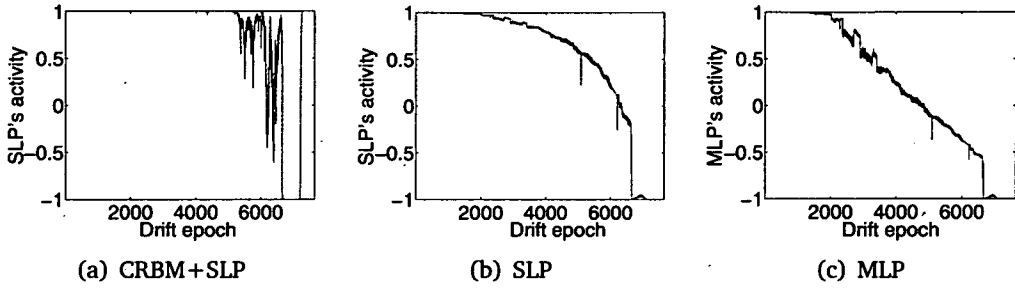
Figure 6.13: *The output of each algorithm over the entire drift period.*

algorithm (i.e. CRBM + SLP, SLP and MLP) were also recorded over the 7644 drift epochs to evaluate the performance of each algorithm. Since merely data from Class A was provided, the output should always be "1" ideally. Fig.6.13 shows the response at the output stage for each algorithm. The neural system with its online updating outperformed the other algorithms as expected. It managed to stay at "1" for the first 5000 epochs. However, it decreased to lower values as the drift rate increased. This was due to the use of fixed learning rate. The limited weight update step restrained the model from adapting to the required amount of change. If the drift rate was continually increased, the difference between the neural system and the actual sensory model outputs would increase, and the neural system would eventually fail completely as seen at the right end of Fig.6.13(a). Meanwhile the non-adaptive models had their responses deteriorated inevitably with the increasing, temporal drift. In comparison to the SLP, the MLP degraded in a steeper slope, probably because of its better-defined, discriminant hyperplane which made it more susceptible to its inputs.

The above simulations have identified that a **fixed learning rate** η_w could be potentially problematic for dynamic environments with varying drift rates as commonly found in most real world applications. If $\eta_w \ll \eta_{op}$ where η_{op} is the optimum learning rate, the model will lose track of the drift eventually, causing a major issue for long term measurements in particularly. On the other hand, if $\eta_w \gg \eta_{op}$, the relatively large weight update step size, results in a model which is very susceptible to noise and can set the model into oscillation around the optimum solution. In the worst case scenario, the over-sized weight update can send the weights straight into saturation, resulting in immediate failure of data classification.

Therefore, a series of simulations was carried out to demonstrate that different drift rates have different η_{op}. The above simulation with the neural system was repeated with different η_w ranging from 0.001 to 0.100. The simulated drift ranges from $-0.8\mu V/s$ to $-4.8\mu V/s$. In each simulation, the trained neural system was fed with data drifting at a specific rate, and the model was updated with a learning rate η_w. The drift rate was defined by the total drift over the first 6000 epochs. Fig.6.14 depicts the classification accuracy of the model at the 6000th drift epoch with different learning and drift rates. Several observations on the figure were made:

- At low drift rate ($-0.8\mu V/s$), the classification accuracy deteriorated with increasing learning rates. The effect was more glaring than of higher drift rates. This is due to the oversized weight updates for a relatively gentle but noisy drift.

- With a fixed learning rate of 0.001, the classification accuracy degraded with the increase in drift rate. Higher learning rates were required to improve the accuracy for faster drifting data. For example, the optimum learning rate was 0.0075 at drift rate of $-4.8\mu V/s$.

- Balancing between faster response and robustness against noise is non-trivial in the online learning. In contrast to the 100% accuracy achieved at drift and learning rates of $-1.6\mu V/s$ and 0.001, the adaptive model could only score 90.82% in accuracy at drift and learning rates of $-4.8\mu V/s$ and 0.0075 respectively.

6.4 Discussion

Potentiostatic techniques used in pH sensing are applicable to many other electrochemical analyses. Example applications include electronic tongues and noses [153–156]. While the sensor fabrication techniques are now well-established, the long-term stability in reference electrodes remains a challenge. Inevitably, these electrochemical sensors suffer from stochastic drift and therefore require constant autocalibration.

In this chapter, the proposed neural system and its learning methodology were examined in two different trials. The first trial was about high-dimensional, simple yet overlapping data classification problem in a dynamic environment. The overlapping was due to limited sensivity and noise induced by interfering ions. Four reference electrodes
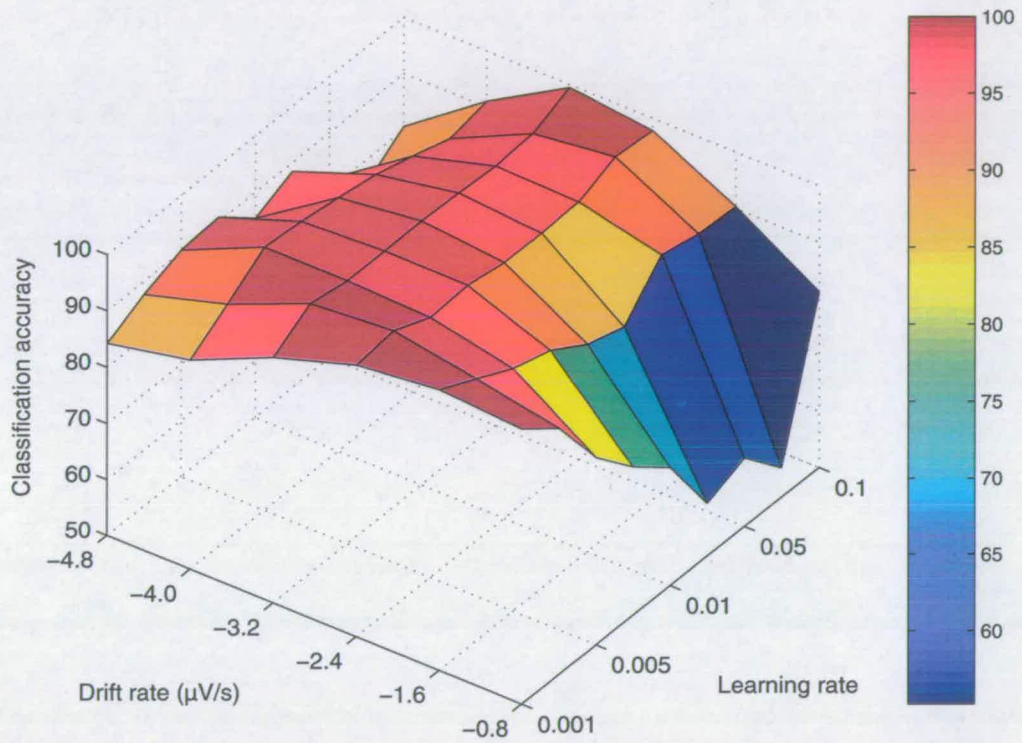
Figure 6.14: *The classification accuracy at different drift and learning rates. With pH-ISFET sensivity of -23mV/pH, drift rate of $-1\mu V/s$ is equal to $+42.735\mu pH/s$.*

were used to improve the robustness of the system against sensor drift and malfunction. While all the pH-ISFET sensors experienced drift at different rates, the neural system managed to stay 100% accurate in test data classification for the first 6000 drift epochs. Two trained but subsequently non-adaptive neural models, namely SLP and MLP, were employed as benchmarks to elucidate the importance of online learning. Both models experienced degration in their classification accuracy with the increasing sensor drift. By the 6000th epoch, their classifications were less than 58% accurate.

In addition, the neural system showed its robustness against sensor failure. Such failure was caused by poor sealing of the microreservoir of the reference electrode and occured more frequently in harsh environment. Traditionally, a median filter or a fault detector is required to validate or/and isolate the erroneous measurements [157–159]. In this work, the probability of the unknown input data being Class A/B was calculated based on the weighted current states of *all* sensors, arguably a variant of median filters. The model would not, understandably, be able to cope with simultaneous malfunction in large number of sensors.

The second trial was a more difficult data modelling problem - meshed clusters. A simple linear classifier (i.e. SLP) could only achieve 95.2% in accuracy even after the lengthy 20000 training epochs. Fast drifting data (hence larger drift step) was used to get a clearer result in the test. The neural system performance in data classification was compared with the two benchmarks at three time intervals. The simulation results as listed in Table 6.2 proved that (1) the neural system was also applicable to non-Gaussian, non-linear classification problems, and (2) it could out-perform once again the non-adaptive SLP and 7:7:1 MLP. The benchmarks had their boundary planes fixed after their initial training stage. Hence, in the dynamic environment, it was not surprising that the augmenting drift increased the number of misclassification.

Admittedly, a MLP with a complex compensation scheme may be adaptive to the sensor drift. In [151], the causes of drift are known but their relationship to drift are not clear. Therefore, a MLP is used to map the relationship between the causes (i.e. temperature and pressure) and the sensor drift, as a compensation scheme. Note that, the MLP is pre-trained with *all possible scenarios* which the sensors will experience, that is not always available in most real-world applications. In the absence of such information, unsupervised neural algorithms are usually the more favourable options than the su-

pervised ones. Furthermore, the drift scenario examined in this chapter is non-trivial. The neural system has to learn the drift from one class of data only over a long period of time. It is anticipated that models such as Kohonen's SOM will experience CI. This will be confirmed in future simulation work.

From the simulation results, it becomes apparent that the learning rate, which determines the online weight update step size, has a delicate role in ensuring proper adaptation is applied. Too high a rate sends the model into oscillation and possibly saturation; too low a rate constraints the adaptation and potentially causes the model to lose track of the drift. This is a common problem in online learning when it is applied in a changing environment. One possible solution is to use an adaptive learning rate. In [160], Amari first proposed the concept of *learning of the learning rule*. The learning rate is trained such that it will be increased to speed up the convergence rate when its weights are far from optimal, and be decreased to improve its classification accuracy when they are near optimal. While these properties are ideal, the rule is only applicable for distance loss function as the cost function of the model to guarantee a convergence and no local minima is present. A more general learning algorithm was then proposed by Murata *et al*[161] which assumed no explicit cost function. This algorithm could potentially be applicable to the neural system.

6.5 Summary

This chapter first illustrated the experiment conducted for obtaining samples of drifting data on a test bench. The data was then used as a representation of a changing environment. It was fused through a trained neural system (CRBM + SLP) in two separate trials: (a) simple, high-dimensional and overlapping, and (b) 2-dimensional, non-linear binary classification problems. During the drift period, the system was allowed to auto-calibrate itself using the training algorithm and the methodology proposed in Chapter 4. Importantly, the learning rate for the CRBM was carefully chosen to match the drift rate. The simulation results proved that the neural system was effective in tracking the imposed sensor drift. This is particularly intriguing because it demonstrates that the CRBM can be configured to respond sensibly to incomplete and "unbalanced" real-time input data that do not adequately represent the distributions of the training data.

Two trained but subsequently non-adaptive benchmarks, namely SLP and MLP, were used to highlight the importance of the online learning to compensate drift. As expected, both algorithms were outperformed by the neural system in both trials. Additionally, the neural system has displayed great robustness against unpredictable sensor failure in Section 6.2 and the experiment omitted in this thesis but can be found in [162]. Under similar conditions, the SLP and the MLP failed immediately. On the contrary, the neural system together with the proposed updating rule managed to recover from the sensor failure(s) within a short period of time. This is particularly useful and important for the electrochemical sensing applications that form the motivation for this study, where robustness in the face of miniature, noisy and unreliable sensors is the primary challenge.

Additionally, the learning rate has been identified as an important determinant of the success on drift tracking. A series of simulations on various drift and learning rates was carried out. The results showed that the optimum learning rate varied proportionally with the drift rate. A compelling solution is to employ an adaptive learning rate. The derivation of its learning rule will form part of the future work for this thesis.

Chapter 7
Summary and Conclusions

In this chapter, a summary of the work done in this thesis is first presented. Subsequently, some conclusions are drawn from the work, with a revisit to the original objective of this thesis. The chapter ends with a list of possibilities of future work.

7.1 Summary

Fig.7.1 illustrates the research progress in this thesis, starting with some literature reviews (Chapters 1 & 2), then some preparatory works (Chapters 3 & 4) and finally software simulations with real and artificial datasets (Chapters 5 & 6).

Firstly, the literature review showed that there was a recent technological trend where more redundant sensors were being implemented on microsystems as the sensor fabrication techniques advanced. These microsystems were deployed in an emerging technology - *distributed sensor networks*. With limited bandwidth in communications, these distributed microsystems must be able to perform local sensor fusion. The project *'IDEAS'* was introduced and the specifications for the neural system were listed. Subsequently, three disciplines directly relevant to this study, namely intelligent sensor fusion in electrochemical sensing, counter-drift techniques and neural hardware in aVLSI were reviewed.

Before the examination of the proposed 'neural' solution (i.e. CRBM) as an effective sensor fusion mechanism, the sensor models and the neural system needed to be developed.

- **Sensor models** - The motivation to derive sensor models was because real sensor data was usually too expensive to be obtained and might be insufficient for training the CRBM. Therefore, significant amount of time was spent on the development and the testing of the control chips. Sensor models were derived from

104
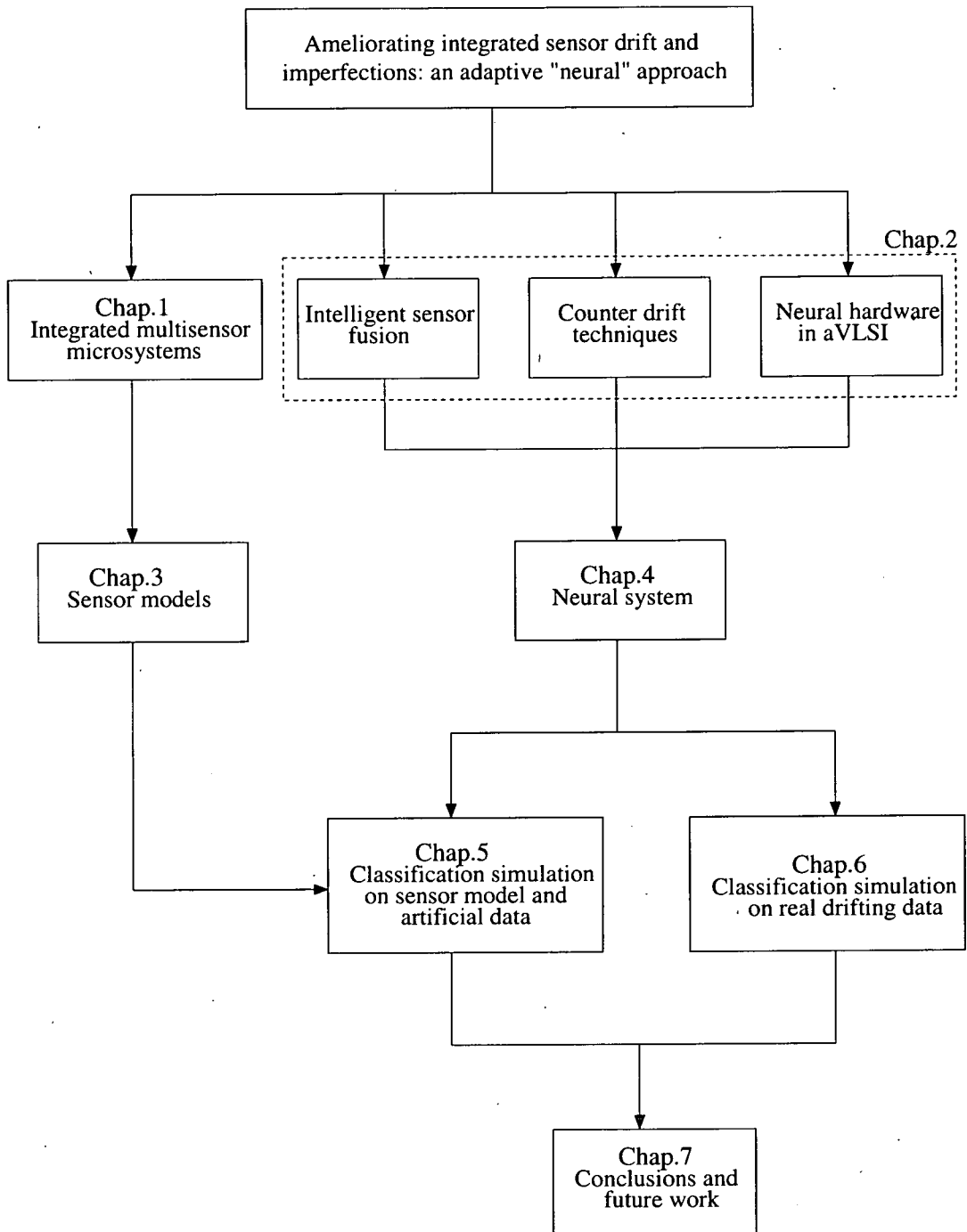
Figure 7.1: *Flowchart illustrating the research progress in this thesis.*

the calibrations on each sensor. Problems such as the short lifespan of a reference electrode and the double layer capacitance effect were experienced in the dissolved oxygen and the conductivity sensors respectively. Only the temperature and the pH sensors were suitable for subsequent simulations. While attempting to excite pH sensors with more current, it was observed that some sensors gave non-linear response due to their wide ranging intrinsic threshold voltages.

- **Neural system** - The neural system was formed by a CRBM for feature extraction and a SLP as a linear classifier. A unique training methodology was proposed and outlined. "Dual-memory" configuration was identified as a way that (i) permitted a permanent encoding of the discriminating feature and (ii) allowed the CRBM being adaptive to a changing environment. During the development, a new online weight learning rule for the CRBM was formulated. This rule ensured a consistent set of features being presented to the non-adaptive SLP. It was argued that this new rule could be easily implemented based on the existing CRBM hardware.

This thesis has two main simulations, namely 'modelling' and 'dynamic adaptation' simulations.

- **On modelling** - The objective was to examine if the neural system could model complex, high-dimensional data distributions. The term 'complex' referred to situations where clusters had meshed areas and were impossible to be separated linearly. The simulation was divided into two parts: part I examined simple but overlapping high-dimensional Gaussian clusters while part II examined 2D meshed non-Gaussian clusters. In part I, 100% accuracy in classification was obtained despite the overlapping between clusters. When input dimension (i.e. the number of sensors) increases, there was more information available and unsurprisingly the neural system became more robust. In part II, the learning was more difficult and few experts (hidden neurons) were needed to represent one non-Gaussian distribution. Longer training period and slower learning rate were required to capture the fine details and to escape from local minima while in search of the global convergence point. Almost 100% accuracy was recorded in two separate trials. Additionally, two parameters in the CRBM learning - learning rate and number of hidden neurons - were also studied. The simulation result agreed with other works in the literature that a slow learning rate could provide a slow

but stable convergence while a fast learning rate resulted a fast but noisy convergence (due to over-sized learning steps). The other simulation result showed that the minimum number of hidden units required to encode the distributions in Fig.5.12(a) was five.

- **On dynamic adaptation** - The objective was to examine if the neural system could adapt to sensor drift. Experiments were carried out to obtain real drift data, and were used in two trials. *Trial 1* had two high-dimensional, overlapping Gaussian clusters; two: *Trial 2* had two 2D meshed, non-Gaussian clusters. Their online adaptation performance was compared with two benchmarks - trained but subsequently non-adaptive SLP and MLP. In *Trial 1*, several different drift and failure rates were experienced in different sensors. The neural system managed to compensate the drift by adjusting its weights in the CRBM. This permitted it to sustain high accuracy rate in classification over longer drift period than the two benchmarks. Similar results were observed in *Trial 2*. Finally, a series of simulations have been run to study the relationship between the drift and the learning rates.

Based on the forementioned simulation results, some conclusions can be drawn and they are described in the following section.

7.2 Conclusions

This thesis examined the suggestion that

> *"local pre-processing and early classification of real, high-dimensional, drifting sensory signals can be achieved effectively by a hardware-amenable neural system"*.

The literature review highlighted an emerging technology - *distributed sensor network*, which has a growing demand in ubiquitous processors to handle local measurement results. In this particular case, 'neural' approach as a form of intelligent sensor fusion is of interest. There is a long list of available neural networks but few are suitable for the application (i.e. the *IDEAS* project), where (a) drift is inevitable, (b) no reference

buffers are available for recalibration, (c) occlusion on sensing surface is common, (d) low power and small silicon footprint of neural hardware implementation are essential. The ideal solution is an unsupervised, adaptive (i.e. on-chip learning) and probabilistic neural algorithm which works with analogue sensory signals directly and is hardware-amenable. These point to CRBM as the most suitable solution.

To ease the binary classification (the sensor fusion mechanism in the *IDEAS* project), the most straightforward approach is to maximise the separation between the two clusters of data. From a circuit design point of view, this involves increasing the sensitivity of the sensors. Two versions of signal conditioning circuits for the pH sensors were implemented. They worked as intended. Given that the power rail was restricted to 3V and the intrinsic threshold voltage of the pH sensors varied significantly, there was little that could be done to improve the sensor sensivity, whilst keeping a linear response within their dynamic range. Therefore it is compelling to further increase the number of redundant sensors to gain more information on the entity of interest.

The neural system is formed by a CRBM and a SLP. The former extracts the important features from the training data distributions while the latter performs a binary classification. Their implementation is already available in aVLSI technology. However, it is possible for the neural system to experience Catastrophic Interference because, quite often, only a single class of data is available over a lengthy period of time. Therefore, a new weight updating rule for the CRBM during operation in a dynamic environment was proposed. This rule involves only simple operations (i.e. subtraction and multiplication), hence it should be equally hardware-amenable like other parts of CRBM which have been realised in aVLSI technology.

From the 'modelling' simulations (Chapter 5), the neural system has demonstrated that the CRBM has strong modelling capability. Together with a SLP, it achieved 100% accuracy in binary classification in simulations involving (i) high-dimensional, overlapping data distributions and (ii) two-dimensional, complex (with meshed areas) non-Gaussian data distributions. In particular, the latter simulation demanded a non-linear model. A linear model such as SLP could not classify completely-correct.

The selection of two parameters - learning rate and number of hidden units of the CRBM - was also studied. The aim was to find (i) the optimum learning rate which could

ensure a convergence rate and stability, and (ii) the number of hidden units to keep the network size at a minimum. In this particular study (training data as in Fig.5.12(a)), the optimum learning rate was 0.05 and the minimum number of hidden units was five. These parameters are acknowledgedly case dependent, and are determined by the complexity of the training data distributions.

While the CRBM has demonstrated a strong modelling capability, it is not without any weakness. Unlike the original Boltzmann Machine, the CRBM has the noise control factors for each weight. During the simulations, they tended to compete with the weights to encode features in the training data distributions. With more variables such as these in use, training a CRBM became more complicated and tricky. A way to solve this "problem" is to assert some constraints in the learning of the variables. For instance, one can lower the learning rate for the noise control factors of hidden layer, hence allowing other variables (e.g. the noise control factors of visible layer and the weights) to store most key features of the training data distributions first. A further study in managing the learning of these variables is needed to gain more insight into the relationship between the variables.

From the 'dynamic adaptation' simulations (Chapter 6), the neural system has demonstrated its ability to adapt to changing environment. It was examined in two separate trials: (i) high-dimensional, simple data distributions and (ii) two-dimensional, meshed data distributions. The neural system showed its ability to compensate sensor drift by adjusting its weight using the proposed learning rule. This online updating allowed consistent features being presented to the SLP (a binary classifier with fixed weights). To highlight this adaptation feature, it was compared with two trained but subsequently non-adaptive neural models (SLP and MLP). In both trials, the neural models outperformed these two benchmarks. In particular, the neural system demonstrated a good robustness against sensor drift at various rates and one sensor failure, making full use of multiple reference electrodes (assuming that they experienced different drift and malfunction rates).

Moreover, *online learning rate* in CRBM was identified as a determinant factor in the autocalibration process. As shown in Fig.6.14, different drift rates require different learning rates. If the learning rate is set to be higher than the optimum one, the CRBM will be more responsive to stochastic noise and possibly send its weights into saturation.

On the other hand, if the learning rate is set to be lower than the optimum one, the CRBM will slowly but surely lose track of the drift eventually. As drift rate is likely to vary with time, using a *fixed* learning rate is not ideal. One solution is to use an *adaptive* learning rate.

Supported by the evidence shown in the 'modelling' and 'dynamic adaptation' simulation results, it is finally concluded that the neural system can be an answer to effective data fusion in distributed sensor networks, albeit with some weaknesses. Improving these weaknesses will be the main task of future work.

7.3 Future work

The simulations reveal that there is a lack of full understanding on the relationship between weights and noise control factors. Inadequate training methodology results in a competition among the variables to encode features in the training distributions. The neural system relies wholly on the bias units in the CRBM to store the baseline of the training distributions and the thresholds for the hidden layer, and updates only the weights of these bias units in the dynamic environment. However, if a feature is represented by a collection of variables, the proposed autocalibration approach may be less effective. A further study on the variables will provide some useful insight on how to derive a better training methodology.

Additionally, the simulations manifest a demand of an adaptive learning rate for the neural system to operate more efficiently in a dynamic environment. An oversized learning rate results a noise-sensitive neural system while an undersized learning rate poses a danger in the CRBM losing track of sensor drift. A general adaptive learning rule proposed by Murata [161] is potentially applicable to the CRBM. For a start, one needs to integrate the MCD learning rule for the weights into the learning rule of the learning rate. Simultaneously, the appropriate values for the algorithm parameters have to be determined to ensure stability in the learning.

The natural next step will be the implementation of the complete neural system in aVLSI technology. Issues such as data communication method and memory for short/long term will require some considerations. The neural hardware can then be used as a plat-form, together with a MCU/DSP that has similar ISF functions, to perform a compar-

ative study of different architectural processors in terms of efficiency in performance, power and silicon area. The ultimate goal is to exploit the parallelism in the neural hardware, an area that recently attracts much research interest under a discipline, named "neuromorphic engineering".

Subsequently, the neural hardware should be integrated with multisensor microsystems in a distributed network. The neural system has a probabilistic behaviour. Therefore, further exploration for applications is required to highlight its robustness against measurement errors caused by occasional/permanent occlusion on sensing surface etc. In particular, it is anticipated that such a neural system will flourish in applications where the key features to distinguish different data clusters are encoded in the correlation between different spatial measurements.

In this thesis, merely simple, uncompensated MLP was used as a benchmark in the 'dynamic simulation' simulations. It will be of interest to implement a more complex MLP to examine the strengths and the weaknesses of both supervised and unsupervised neural algorithms in the forementioned application where no control measurement is available. On unsupervised neural algorithms, a further comparison between the proposed neural system and the Kohonen's SOM will be assistive to highlight the strength of the proposed neural system in real-world applications where merely one class of data is present over a long measurement period.

Another possible future work is to analyse the capabilities of CRBM. Such analysis will be useful to find the limitations of the CRBM and hence inspire further development in it. A fully mathematical analysis may be very difficult. Extensive simulations are useful alternatives.

Appendix A
List of publications

The work from this thesis has generated several publications. They are listed below.

A.1 International Refereed Journal Papers

1. E.A. Johannessen, L. Wang, L. Cui, T.B. Tang, M. Ahmadian, A. Astaras, S.W. Reid, P. Yam, A.F. Murray, B. W. Flynn, S.P. Beaumont, D.R.S. Cumming, J.M. Cooper, "Implementation of Distributed Sensors in a Microsystems Format", *IEEE Transactions on Biomedical Engineering*, vol 51(3), March 2004, pp 525-535.

2. T.B. Tang, H. Chen, A.F. Murray, "Adaptive, Integrated Sensor Processing to Compensate for Drift and Uncertainty: A Stochastic "Neural" Approach", *IEE Proceedings on Nanobiotechnology*, vol 151(1), February 2004, pp 28-34.

3. T.B.Tang, E. Johannessen, L. Wang, A. Astaras, M. Ahmadian, A.F. Murray, J.M. Cooper, S.P. Beaumont, B.W. Flynn, D.R.S. Cumming, "Toward a Miniature Wireless Integrated Multisensor Microsystem for Industrial and Biomedical Applications", *IEEE Sensors Journal: Special Issue on Integrated Multisensor Systems and Signal Processing*, vol 2(6), December 2002, pp 628-635.

A.2 International Conference Papers/Abstracts

1. T.B. Tang, A.F. Murray, "Adaptive Sensor Modelling and Classification using a Continuous Restricted Boltzmann Machine (CRBM)", *Proc. of the 14th European Symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium, April, 2006, accepted.

2. A. Astaras, T.B. Tang, L. Wang, A.F. Murray, S.P. Beaumont, D.R.S. Cumming, "Noise Analysis on Integrated Multisensor Microsystems", *Proc. of the 3rd IEEE Sensors Conference*, Vienna, Austria, October 2004, pp 995-998.

3. L. Wang, P. Hammond, E. Johannessen, T.B. Tang, A. Astaras, S.P. Beaumont, A.F. Murray, J.M. Cooper, D.R.S. Cumming, "An On-Chip Programmable Instrumentation Microsystem for Gastrointestinal Telemetry Applications", *Proc. of the 26th Annual International Conference of the IEEE EMBS*, San Francisco, CA, USA, September 2004, pp 2109-2112.

4. T.B. Tang, H. Chen, A.F. Murray, "Adaptive Stochastic Classifier for Noisy pH-ISFET Measurement", *Proc. of the 13th International Conference on Artificial Neural Networks (ICANN'03)*, Istanbul, Turkey, Springer-Verlag, June 2003, pp 638-645.

5. L. Wang, E.A. Johannessen, L. Cui, Colin Ramsay, T.B. Tang, M. Ahmadian, A. Astaras, Peter W. Dickman, J.M. Cooper, A.F. Murray, B.W. Flynn, S.P. Beaumont, D.R.S. Cumming, "Networked Wireless Microsystem for Remote Gastrointestinal Monitoring", *Digest of technical papers for the 12th International Conference on Solid-State Sensors, Actuators and Microsystems*, Boston, USA, June 2003, pp 1184-1187.

6. H. Chen, P. Fleury, T.B. Tang, A.F. Murray, "Adaptive Noisy Neural Computation in Mixed-mode VLSI", *Proc. of the Seventh International Conference on Cognitive and Neural Systems*, Boston, USA, May 2003, pp 68.

7. A. Astaras, M. Ahmadian, N. Aydin, L. Cui, E. Johannessen, T.B. Tang, L. Wang, T. Arslan, S.P. Beaumont, B.W. Flynn, A.F. Murray, S.W. Reid, P. Yam, J.M. Cooper, D.R.S. Cumming, "A miniature integrated electronics sensor capsule for real-time monitoring of the gastrointestinal tract (IDEAS)", *Proc. of the IEEE ICBME Conference*, Singapore, December 2002.

8. T.B. Tang, E. Johannessen, L. Wang, A. Astaras, M. Ahmadian, L. Cui, A.F. Murray, J.M. Cooper, S.P. Beaumont, B.W. Flynn, D.R.S. Cumming, "IDEAS: A Miniature Lab-in-a-Pill Multisensor Microsystem", *Proc. of the IEEE Norchip Conference*, Copenhagen, Denmark, November 2002, pp 329-334.

9. E. A. Johannessen, T. B. Tang, L. Wang, L. Cui, M. Ahmadian, N. Aydin, A. Astaras, A. F. Murray, B. W. Flynn, T. Arslan, S. P. Beaumont, D. R. S. Cumming, J. M. Cooper, "An ingestible electronic pill for real time analytical measurements of the gastrointestinal tract", *Proc. of the uTAS 2002 Symposium*, Japan, November 2002, pp 181-183.

113

10. L. Wang, T.B. Tang, E. Johannessen, A. Astaras, A.F. Murray, J.M. Cooper, S.P. Beaumont, D.R. S. Cumming, "An integrated sensor microsystem for industrial and biomedical applications", *Proc. of the IEEE-EMBS Special Topic Conference on Microtechnologies in Medicine & Biology, Madison*, Wisconsin, USA, May 2002, pp 219-222.

11. L. Wang, T.B. Tang, E. Johannessen, A. Astaras, M. Ahmadian, A.F. Murray, J.M. Cooper, S.P. Beaumont, D.R. S. Cumming, "Integrated micro-instrumentation for dynamic gastro-intestinal tract monitoring", *Proc. of the IEEE IMTC*, Anchorage, Alaska, USA, May 2002, pp 1717-1720.

Implementation of Multichannel Sensors for Remote Biomedical Measurements in a Microsystems Format

Erik A. Johannessen*, Lei Wang, *Member, IEEE*, Li Cui, Tong Boon Tang, *Student Member, IEEE*, Mansour Ahmadian, Alexander Astaras, *Student Member, IEEE*, Stuart W. J. Reid, Philippa S. Yam, Alan F. Murray, *Senior Member, IEEE*, Brian W. Flynn, Steve P. Beaumont, David R. S. Cumming, *Member, IEEE*, and Jonathan M. Cooper

Abstract—A novel microelectronic "pill" has been developed for *in situ* studies of the gastro-intestinal tract, combining microsensors and integrated circuits with system-level integration technology. The measurement parameters include real-time remote recording of temperature, pH, conductivity, and dissolved oxygen. The unit comprises an outer biocompatible capsule encasing four microsensors, a control chip, a discrete component radio transmitter, and two silver oxide cells (the latter providing an operating time of 40 h at the rated power consumption of 12.1 mW). The sensors were fabricated on two separate silicon chips located at the front end of the capsule. The robust nature of the pill makes it adaptable for use in a variety of environments related to biomedical and industrial applications.

Index Terms—Microelectronic pill, microsensor integration, mobile analytical microsystem, multilayer silicon fabrication, radiotelemetry, remote *in situ* measurements.

I. INTRODUCTION

THE invention of the transistor enabled the first radiotelemetry capsules, which utilized simple circuits for *in vivo* telemetric studies of the gastro-intestinal (GI) tract [1]. These units could only transmit from a single sensor channel, and were difficult to assemble due to the use of discrete components [2]. The measurement parameters consisted of either temperature, pH or pressure, and the first attempts of conducting real-time noninvasive physiological measurements suffered from poor reliability, low sensitivity, and short lifetimes of the devices. The first successful pH gut profiles were achieved in 1972 [3], with subsequent improvements in

sensitivity and lifetime [4], [5]. Single-channel radiotelemetry capsules have since been applied for the detection of disease and abnormalities in the GI tract [6]–[8] where restricted access prevents the use of traditional endoscopy [9].

Most radiotelemetry capsules utilize laboratory type sensors such as glass pH electrodes, resistance thermometers [10], or moving inductive coils as pressure transducers [11]. The relatively large size of these sensors limits the functional complexity of the pill for a given size of capsule. Adapting existing semiconductor fabrication technologies to sensor development [12]–[17] has enabled the production of highly functional units for data collection, while the exploitation of integrated circuitry for sensor control, signal conditioning, and wireless transmission [18], [19] has extended the concept of single-channel radiotelemetry to remote distributed sensing from microelectronic pills.

Our current research on sensor integration and onboard data processing has, therefore, focused on the development of microsystems capable of performing simultaneous multiparameter physiological analysis. The technology has a range of applications in the detection of disease and abnormalities in medical research. The overall aim has been to deliver enhanced functionality, reduced size and power consumption, through system-level integration on a common integrated circuit platform comprising sensors, analog and digital signal processing, and signal transmission.

In this paper, we present a novel analytical microsystem which incorporates a four-channel microsensor array for real-time determination of temperature, pH, conductivity and oxygen. The sensors were fabricated using electron beam and photolithographic pattern integration, and were controlled by an application specific integrated circuit (ASIC), which sampled the data with 10-bit resolution prior to communication off chip as a single interleaved data stream. An integrated radio transmitter sends the signal to a local receiver (base station), prior to data acquisition on a computer. Real-time wireless data transmission is presented from a model *in vitro* experimental setup, for the first time.

Details of the sensors are provided in more detail later, but included: a silicon diode [20] to measure the body core temperature, while also compensating for temperature induced signal changes in the other sensors; an ion-selective field effect transistor, ISFET, [21] to measure pH; a pair of direct contact gold electrodes to measure conductivity; and a three-electrode electrochemical cell [22], to detect the level of dissolved oxygen in solution. All of these measurements will, in the future, be used to perform *in vivo* physiological analysis of the GI-tract.

For example, temperature sensors will not only be used to measure changes in the body core temperature, but may also identify local changes associated with tissue inflammation and ulcers. Likewise, the pH sensor may be used for the determination of the presence of pathological conditions associated with abnormal pH levels, particularly those associated with pancreatic disease and hypertension, inflammatory bowel disease, the activity of fermenting bacteria, the level of acid excretion, reflux to the oesophagus, and the effect of GI specific drugs on target organs. The conductivity sensor will be used to monitor the contents of the GI tract by measuring water and salt absorption, bile secretion and the breakdown of organic components into charged colloids. Finally, the oxygen sensor will measure the oxygen gradient from the proximal to the distal GI tract. This will, in future enable a variety of syndromes to be investigated including the growth of aerobic bacteria or bacterial infection concomitant with low oxygen tension [23], as well as the role of oxygen in the formation of radicals causing cellular injury and pathophysiological conditions (inflammation and gastric ulceration). The implementation of a generic oxygen sensor will also enable the development of first generation enzyme linked amperometric biosensors, thus greatly extending the range of future applications to include, e.g., glucose and lactate sensing, as well as immunosensing protocols.

II. MICROELECTRONIC PILL DESIGN AND FABRICATION

A. Sensors

The sensors were fabricated on two silicon chips located at the front end of the capsule. *Chip 1* [Fig. 1(a), (c), (e)] comprises the silicon diode temperature sensor, the pH ISFET sensor and a two electrode conductivity sensor. *Chip 2* [Fig. 1(b), (d), (f)] comprises the oxygen sensor and an optional nickel-chromium (NiCr) resistance thermometer. The silicon platform of *Chip 1* was based on a research product from Ecole Superieure D'Ingenieurs en Electrotechnique et Electronique (ESIEE, France) with predefined n-channels in the p-type bulk silicon forming the basis for the diode and the ISFET. A total of 542 of such devices were batch fabricated onto a single 4-in wafer. In contrast, *Chip 2* was batch fabricated as a 9×9 array on a 380-μm-thick single crystalline 3″ silicon wafer with $\langle 100 \rangle$ lattice orientation, precoated with 300 nm Si_3N_4, silicon nitride, (Edinburgh Microfabrication Facility, U.K.). One wafer yielded 80, 5×5 mm^2 sensors (the center of the wafer was used for alignment markers).

1) Sensor Chip 1: An array of 4×2 combined temperature and pH sensor platforms were cut from the wafer and attached on to a 100-μm-thick glass cover slip using S1818 photoresist (Microposit, U.K.) cured on a hotplate. The cover slip acted as temporary carrier to assist handling of the device during the first level of lithography (*Level 1*) when the electric connection tracks, the electrodes and the bonding pads were defined. The pattern was defined in S1818 resist by photolithography prior to thermal evaporation of 200 nm gold (including an adhesion layer of 15 nm titanium and 15 nm palladium). An additional layer of gold (40 nm) was sputtered to improve the adhesion of the electroplated silver used in the reference electrode (see below). Liftoff in acetone detached the chip array from the cover slip. Individual sensors were then diced prior to their re-attachment in pairs on a 100-μm-thick cover slip by epoxy resin
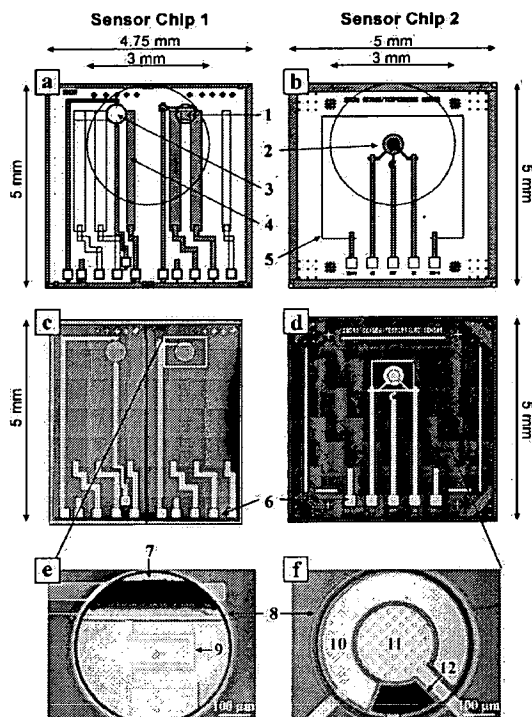
Fig. 1. The microelectronic sensors: (a) schematic diagram of *Chip 1*, measuring 4.75×5 mm^2, comprising the pH (ISFET) sensor (1), the 5×10^{-4} mm^2 dual electrode conductivity sensor (3) and the silicon diode temperature sensor (4); (b) schematic diagram of *Chip 2*, measuring 5×5 mm^2, comprising the electrochemical oxygen sensor (2) and a NiCr resistance thermometer (5). Once integrated in the pill, the area exposed to the external environment is illustrated by the 3-mm-diameter circle; (c) photomicrograph of sensor *Chip 1* and (d) sensor *Chip 2*. The bonding pads (6), which provide electrical contact to the external electronic control circuit, are shown; (e) close up of the pH sensor consisting of the integrated 3×10^{-2} mm^2 Ag|AgCl reference electrode (7), a 500-μm-diameter and 50-μm-deep, 10-nL, electrolyte chamber (8) defined in polyimide, and the 15×600 μm floating gate (9) of the ISFET sensor; (f) the oxygen sensor is likewise embedded in an electrolyte chamber (8). The three-electrode electrochemical cell comprises the 1×10^{-1} mm^2 counter electrode (10), a microelectrode array of 57×10 μm diameter (4.5×10^3 mm^2) working electrodes (11) defined in 500-nm-thick PECVD Si_3N_4, and an integrated 1.5×10^{-2} mm^2 Ag|AgCl reference electrode (12).

[Fig. 1(c)]. The left-hand-side (LHS) unit comprised the diode, while the right-hand-side (RHS) unit comprised the ISFET. The 15×600 μm (L $\times$ W) floating gate of the ISFET was precovered with a 50-nm-thick proton sensitive layer of Si_3N_4 for pH detection [24].

Photocurable polyimide (Arch Chemicals n.v., Belgium) defined the 10-nL electrolyte chamber for the pH sensor (above the gate) and the open reservoir above the conductivity sensor (*Level 2*).

The silver chloride reference electrode (3×10^{-2} mm^2) was fabricated during *Levels 3* to *5*, inclusive. The glass cover slip, to which the chips were attached, was cut down to the size of the 4.75×5 mm^2 footprint (still acting as a supporting base) prior to attachment on a custom-made chip carrier used for electroplating. Silver (5 μm) was deposited on the gold electrode

defined at by chronopotentiometry (-300 nA, 600 s) after removing residual polyimide in an O_2 barrel asher (Electrotech, U.K.) for 2 min. The electroplating solution consisted of 0.2 M $AgNO_3$, 3 M KI and 0.5 M $Na_2S_2O_3$. Changing the electrolyte solution to 0.1 M KCl at *Level 4* allowed for the electroplated silver to be oxidized to AgCl by chronopoteniometry (300 nA, 300 s). The chip was then removed from the chip carrier prior to injection of the internal 1 M KCl reference electrolyte required for the Ag | AgCl reference electrode (*Level 5*). The electrolyte was retained in a 0.2% gel matrix of calcium alginate [25].

The chip was finally clamped by a 1-mm-thick stainless-steel clamp separated by a 0.8-μm-thick sheet of Viton fluoroelastomer (James Walker, U.K.). The rubber sheet provided a uniform pressure distribution in addition to forming a seal between the sensors and capsule.

2) Sensor Chip 2: The *level 1* pattern (electric tracks, bonding pads, and electrodes) was defined in 0.9 μm UV3 resist (Shipley, U.K.) by electron beam lithography. A layer of 200 nm gold (including an adhesion layer of 15 nm titanium and 15 nm palladium) was deposited by thermal evaporation. The fabrication process was repeated (*Level 2*) to define the 5-μm-wide and 11-mm-long NiCr resistance thermometer made from a 100-nm-thick layer of NiCr (30-kΩ resistance). *Level 3* defined the 500-nm-thick layer of thermal evaporated silver used to fabricate the reference electrode. An additional sacrificial layer of titanium (20 nm) protected the silver from oxidation in subsequent fabrication levels. The surface area of the reference electrode was 1.5×10^{-2} mm^2, whereas the counter electrode made of gold had an area of 1.0×10^{-1} mm^2.

Level 4 defined the microelectrode array of the working electrode, comprising 57 circular gold electrodes, each 10 μm in diameter, with an interelectrode spacing of 25 μm and a combined area of 4.5×10^{-3} mm^2. Such an array promotes electrode polarization and reduces response time by enhancing transport to the electrode surface [26]. The whole wafer was covered with 500 nm plasma-enhanced chemical vapor deposited (PECVD) Si_3N_4. The pads, counter, reference, and the microelectrode array of the working electrode was exposed using an etching mask of S1818 photoresist prior to dry etching with C_2F_6. The chips were then diced from the wafer and attached to separate 100-μm-thick cover slips by epoxy resin to assist handling. The electrolyte chamber was defined in 50-μm-thick polyimide at *Level 5*. Residual polyimide was removed in an O_2 barrel asher (2 min), prior to removal of the sacrificial titanium layer at *Level 6* in a diluted HF solution (HF to RO water, 1:26) for 15 s. The short exposure to HF prevented damage to the PECVD Si_3N_4 layer.

Thermally evaporated silver was oxidized to Ag |AgCl (50% of film thickness) by chronopotentiometry (120 nA, 300 s) at *Level 7* in the presence of KCl, prior to injection of the internal reference electrolyte at *Level 8*. A 5×5 mm^2 sheet of oxygen permeable teflon was cut out from a 12.5-μm-thick film and attached to the chip at *Level 9* with epoxy resin prior to immobilization by the aid of a stainless steel clamp.

B. Control Chip

The ASIC was a control unit that connected together the external components of the microsystem (Fig. 2). It was fabricated as a 22.5 mm^2 silicon die using a 3-V, 2-poly, 3-metal 0.6-μm
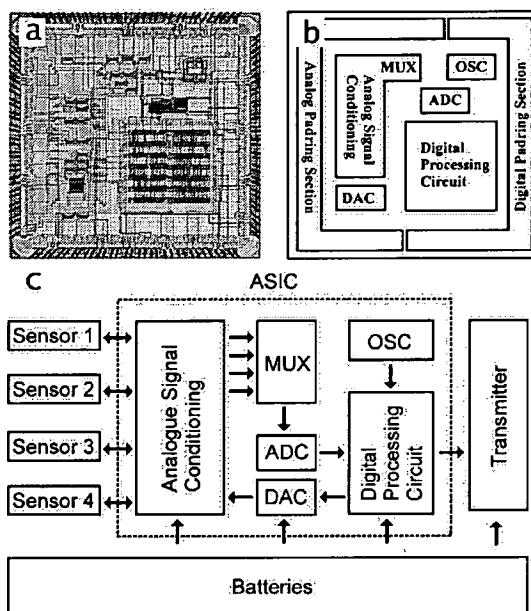
Fig. 2. Photograph of the 4.75×4.75 mm^2 application specific integrated circuit control chip (a), the associated explanatory diagram (b), and a schematic of the architecture (c) illustrating the interface to external components. MUX (four-channel multiplexer), ADC, DAC, and OSC (32-kHz oscillator).

CMOS process by Austria Microsystems (AMS) via the Europractice initiative. It is a novel mixed signal design that contains an analog signal conditioning module operating the sensors, an 10-bit analog-to-digital (ADC) and digital-to-analog (DAC) converters, and a digital data processing module. An *RC* relaxation oscillator (OSC) provides the clock signal.

The analog module was based on the AMS OP05B operational amplifier, which offered a combination of both a power-saving scheme (sleep mode) and a compact integrated circuit design. The temperature circuitry biased the diode at constant current, so that a change in temperature would reflect a corresponding change in the diode voltage. The pH ISFET sensor was biased as a simple source and drain follower at constant current with the drain-source voltage changing with the threshold voltage and pH. The conductivity circuit operated at direct current measuring the resistance across the electrode pair as an inverse function of solution conductivity. An incorporated potentiostat circuit operated the amperometric oxygen sensor with a 10-bit DAC controlling the working electrode potential with respect to the reference. The analog signals had a full-scale dynamic range of 2.8 V (with respect to a 3.1-V supply rail) with the resolution determined by the ADC. The analog signals were sequenced through a multiplexer prior to being digitized by the ADC. The bandwidth for each channel was limited by the sampling interval of 0.2 ms.

The digital data processing module conditioned the digitized signals through the use of a serial bitstream data compression algorithm, which decided when transmission was required by comparing the most recent sample with the previous sampled data. This technique minimizes the transmission length, and

117

is particularly effective when the measuring environment is at quiescent, a condition encountered in many applications [27]. The entire design was constructed with a focus on low power consumption and immunity from noise interference. The digital module was deliberately clocked at 32 kHz and employed a sleep mode to conserve power from the analog module. Separate on-chip power supply trees and pad-ring segments were used for the analog and digital electronics sections in order to discourage noise propagation and interference.

C. Radio Transmitter

The radio transmitter was assembled prior to integration in the capsule using discrete surface mount components on a single-sided printed circuit board (PCB). The footprint of the standard transmitter measured $8 \times 5 \times 3$ mm including the integrated coil (magnetic) antenna. It was designed to operate at a transmission frequency of 40.01 MHz at 20 °C generating a signal of 10 kHz bandwidth. A second crystal stabilized transmitter was also used. This second unit was similar to the free running standard transmitter, apart from having a larger footprint of $10 \times 5 \times 3$ mm, and a transmission frequency limited to 20.08 MHz at 20 °C, due to the crystal used. Pills incorporating the standard transmitter were denoted *Type I*, whereas the pills incorporating the crystal stabilized unit were denoted *Type II*. The transmission range was measured as being 1 meter and the modulation scheme frequency shift keying (FSK), with a data rate of 1 kbs^{-1}.

D. Capsule

The microelectronic pill consisted of a machined biocompatible (noncytotoxic), chemically resistant polyether-terketone (PEEK) capsule (Victrex, U.K.) and a PCB chip carrier acting as a common platform for attachment of the sensors, ASIC, transmitter and the batteries (Fig. 3). The fabricated sensors were each attached by wire bonding to a custom made chip carrier made from a 10-pin, 0.5-mm pitch polyimide ribbon connector. The ribbon connector was, in turn, connected to an industrial standard 10-pin flat cable plug (FCP) socket (Radio Spares, U.K.) attached to the PCB chip carrier of the microelectronic pill, to facilitate rapid replacement of the sensors when required. The PCB chip carrier was made from two standard 1.6-mm-thick fiber glass boards attached back to back by epoxy resin which maximized the distance between the two sensor chips. The sensor chips were connected to both sides of the PCB by separate FCP sockets, with sensor *Chip 1* facing the top face, with *Chip 2* facing down. Thus, the oxygen sensor on *Chip 2* had to be connected to the top face by three 200-μm copper leads soldered on to the board. The transmitter was integrated in the PCB which also incorporated the power supply rails, the connection points to the sensors, as well as the transmitter and the ASIC and the supporting slots for the capsule in which the chip carrier was located.

The ASIC was attached with double-sided copper conducting tape (Agar Scientific, U.K.) prior to wirebonding to the power supply rails, the sensor inputs, and the transmitter (a process which entailed the connection of 64 bonding pads). The unit was powered by two standard 1.55-V SR44 silver oxide (Ag_2O) cells with a capacity of 175 mAh. The batteries were serial connected and attached to a custom made 3-pin, 1.27-mm pitch plug
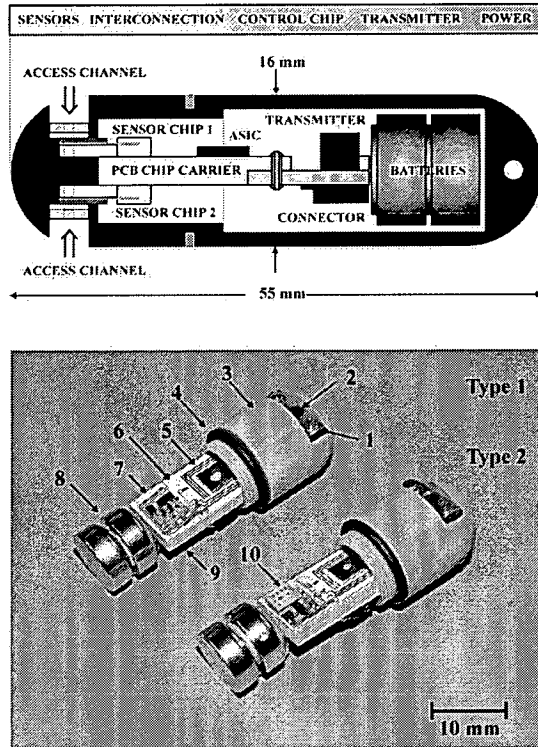
Fig. 3. Schematic diagram (top) of the remote mobile analytical microsystem comprising the electronic pill. The prototype is 16×55 mm, weights 13.5 g. The *Type I* unit consist of the microelectronic sensors at the front enclosed by the metal clamp and rubber seal (1) which provide a 3-mm-diameter access channel to the sensors (2). The front section of the capsule, physically machined from solid PEEK, is illustrated (3) with the rear section removed to illustrate the internal design. The front and rear section of the capsule is joined by a screw connection sealed of by a Viton-rubber o-ring (4). The ASIC control chip (5) is integrated on the common PCB chip carrier (6) which incorporate the discrete component radio transmitter (7), and the silver oxide battery cells (8). The battery is connected on the reverse side of the PCB (9). The *Type II* unit is identical to the *Type I* with exception of an incorporated crystal stabilized radio transmitter (10) for improved temperature stability.

by electrical conducting epoxy (Chemtronics, Kennesaw, GA). The connection to the matching socket on the PCB carrier provided a three point power supply to the circuit comprising a negative supply rail (-1.55 V), virtual ground (0 V), and a positive supply rail (1.55 V). The battery pack was easily replaced during the experimental procedures.

The capsule was machined as two separate screw-fitting compartments. The PCB chip carrier was attached to the front section of the capsule (Fig. 3). The sensor chips were exposed to the ambient environment through access ports and were sealed by two sets of stainless steel clamps incorporating a 0.8-μm-thick sheet of Viton fluoroelastomer seal. A 3-mm-diameter access channel in the center of each of the steel clamps (incl. the seal), exposed the sensing regions of the chips. The rear section of the capsule was attached to the front section by a 13-mm screw connection incorporating a Viton rubber O-ring (James Walker, U.K.). The seals rendered the capsule water proof, as

well as making it easy to maintain (e.g., during sensor and battery replacement). The complete prototype was 16×55 mm and weighted 13.5 g including the batteries. A smaller pill suitable for physiological *in vivo* trials (10×30 mm) is currently being developed from the prototype.

III. MATERIAL AND METHODS

A. General Experimental Setup

All the devices were powered by batteries in order to demonstrate the concept of utilizing the microelectronic pill in remote locations (extending the range of applications from *in vivo* sensing to environmental or industrial monitoring). The pill was submerged in a 250-mL glass bottle located within a 2000-mL beaker to allow for a rapid change of pH and temperature of the solution. A scanning receiver (Winradio Communications, Australia) captured the wireless radio transmitted signal from the microelectronic pill by using a coil antenna wrapped around the 2000-mL polypropylene beaker in which the pill was located. A portable Pentium III computer controlled the data acquisition unit (National Instruments, Austin, TX) which digitally acquired analog data from the scanning receiver prior to recording it on the computer.

The solution volume used in all experiments was 250 mL. The beaker, pill, glass bottle, and antenna were located within a 25×25 cm container of polystyrene, reducing temperature fluctuations from the ambient environment (as might be expected within the GI tract) and as required to maintain a stable transmission frequency. The data was acquired using LabView (National Instruments, Austin, TX) and processed using a MATLAB (Mathworks, Natick, MA) routine.

B. Sensor Characterization

The lifetime of the incorporated Ag | AgCl reference electrodes used in the pH and oxygen sensors was measured with an applied current of 1 pA immersed in a 1.0 M KCl electrolyte solution. The current reflects the bias input current of the operational amplifier in the analog sensor control circuitry to which the electrodes were connected.

The temperature sensor was calibrated with the pill submerged in reverse osmosis (RO) water at different temperatures. The average temperature distribution over 10 min was recorded for each measurement, represented as 9.1 °C, 21.2 °C, 33.5 °C, and 47.9 °C. The system was allowed to temperature equilibrate for 5 min prior to data acquisition. The control readings were performed with a thin wire K-type thermocouple (Radio Spares, U.K.). The signal from the temperature sensor was investigated with respect to supply voltage potential, due to the temperature circuitry being referenced to the negative supply rail. Temperature compensated readings (normalized to 23 °C) were recorded at a supply voltage potential of 3.123, 3.094, 3.071, and 2.983 mV using a direct communication link. Bench testing of the temperature sensor from 0 °C to 70 °C was also performed to investigate the linear response characteristics of the temperature sensor.

The pH sensor of the microelectronic pill was calibrated in standard pH buffers [28] of pH 2, 4, 7, 9, and 13, which reflected the dynamic range of the sensor. The calibration was performed at room temperature (23 °C) over a period of 10 min, with the

pill being washed in RO water between each step. A standard lab pH electrode was used as a reference to monitor the pH of the solutions (Consort n.v., Belgium). The pH channel of the pill was allowed to equilibrate for 5 min prior to starting the data acquisition. Each measurement was performed twice. Bench test measurements from pH 1 to 13 were also performed using an identical control circuit to the ASIC.

The oxygen sensor was bench tested with a standard laboratory potentiostat (Bioanalytical Systems, West Lafayette, IN), over its dynamic range in phospate buffered saline (PBS) using a direct communication link at 23 °C. Cyclic voltammetry with a sweep potential from 0.1 to 0.45 V (versus Ag | AgCl) was performed in 1-mM ferroscene-monocarboxylic acid (FMCA) as a model redox compound, to test the performance of the microelectrode array. A three-point calibration routine was performed at oxygen concentrations of 0 mg L^{-1} (PBS saturated with 2 M Na_2SO_3), 4 mg L^{-1} (PBS titration with 2 M Na_2SO_3) and 8.2 mg L^{-1} (oxygen saturated PBS solution). The solution saturated with dissolved oxygen was equilibrated overnight prior to use. The dissolved oxygen was monitored using a standard Clark O_2 electrode (Orion Research Inc., Beverly, MA). The reduction potential of water was assessed in oxygen depleted PBS, to avoid interference from oxygen, at the same time assessing the lower potential limit that could be used for maximizing the efficiency of the sensor. The voltage was then fixed above this reduction potential to assess the dynamic behavior of the sensor upon injection of saturated Na_2SO_3 in oxygen saturated PBS.

C. Transmission

The pill's transmission frequency was measured with respect to changes in temperature. The *Type I* pill (without crystal) was submerged in RO water at temperatures of 1 °C, 11 °C, 23 °C, and 49 °C, whereas the *Type II* pill (with crystal) was submerged in temperatures of 2 °C, 25 °C, and 45 °C. The change in frequency was measured with the scanning receiver, and the results used to assess the advantage of crystals stabilized units at the cost of a larger physical size of the transmitter.

D. Dynamic Measurements

Dynamic pH measurements were performed with the pill submerged in a PBS solution at 23 °C. The pH was changed from the initial value of 7.3 by the titration of 0.1 M H_2SO_4 and 0.1 M NaOH, respectively. Subsequently, the pH was changed from pH 7.3 to pH 5.5 (after 5 min), pH 3.4 (after 8 min) to pH 9.9 (after 14 min) and back to pH 7.7 (after 21 min). A standard (bench-top) pH electrode monitored the pH of the solution. The solutions were sampled after the pH change, and measured outside the experimental system to prevent electronic noise injection from the pH electrode. The temperature channel was recorded simultaneously.

E. Sensor and Signal Drift

Long term static pH and temperature measurements were performed to assess signal drift and sensor lifetime in physiological electrolyte (0.9% saline) solutions. A temperature of 36.5 °C was achieved using a water bath, with the assay solutions continuously stirred and re-circulated using a peristaltic pump. The sensors were transferred from solutions of pH 4 to pH 7, within 2 h of commencing the experiment, and from pH 7 to pH 10.5,
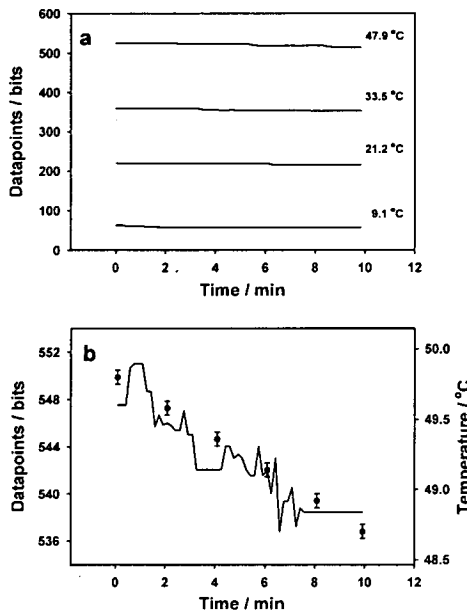
Fig. 4. Temperature sensor: (a) temperature recording over a range from 9.1 °C to 47.9 °C, represented by digital datapoints; (b) high-resolution plot of a temperature change from 49.8 °C to 48.7 °C. The control measurement from the thermocouples is presented as solid points with error bars representing the resolution of the thermometer. The resolution of the temperature channel was noise limited to 0.4 °C.

after 4 h. The total duration of the experiment was 6 h. Each experiment was repeated twice.

IV. RESULTS

The power consumption of the microelectronic pill with the transmitter, ASIC and the sensors connected was calculated to 12.1 mW, corresponding to the measured current consumption of 3.9 mA at 3.1-V supply voltage. The ASIC and sensors consumed 5.3 mW, corresponding to 1.7 mA of current, whereas the free running radio transmitter (*Type I*) consumed 6.8 mW (corresponding to 2.2 mA of current) with the crystal stabilized unit (*Type II*) consuming 2.1 mA. Two SR44 Ag_2O batteries used provided an operating time of more than 40 h for the microsystem.

A. Temperature Channel Performance

The linear sensitivity was measured over a temperature range from 0 °C to 70 °C and found to be 15.4 $mV°C^{-1}$. This amplified signal response was from the analog circuit, which was later implemented in the ASIC. The sensor [Fig. 4(a)], once integrated in the pill, gave a linear regression of 11.9 $bits°C^{-1}$ ($f(T°C) = 11.9 (T°C) - 42.7$, $R^2 = 0.99$), with a resolution limited by the noise band of 0.4 °C [Fig. 4(b)]. The diode was forward biased with a constant current (15 μA) with the n-channel clamped to ground, while the p-channel was floating. Since the bias current supply circuit was clamped to the negative voltage rail, any change in the supply voltage potential would
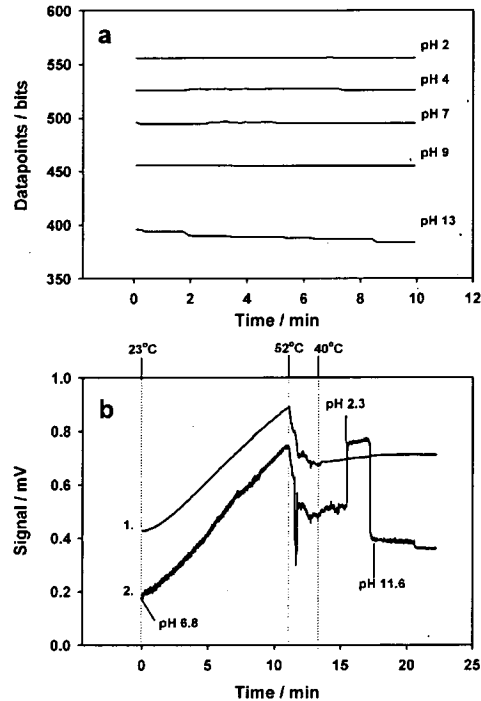
Fig. 5. pH sensor: (a) pH recording in the range of pH 2 to 13, represented by digital datapoints; (b) dynamic recording of temperature (1) and pH (2) using a direct communication link illustrates the temperature sensitivity of the pH channel (16.8 $mV°C^{-1}$), whereas the temperature channel is insensitive to any pH change.

cause the temperature channel to drift. Thus, bench test measurements conducted on the temperature sensor revealed that the output signal changed by 1.45 mV per mV change in supply voltage ($f(mV) = -1.45 (mV) + 2584$, $R^2 = 0.99$) with $f(mV)$ expressed in millivolts, corresponding to a drift of -21 mV h^{-1} in the pill from a supply voltage change of -14.5 mV h^{-1}.

B. pH Channel Performance

The linear characteristics from pH 1 to 13 corresponded to a sensitivity of -41.7 mV pH^{-1} unit at 23 °C, which is in agreement with literature values [21] although the response was lower than the Nernstian characteristics found in standard glass pH electrodes (-59.2 mV pH^{-1} unit). The pH ISFET sensor operated in a constant current mode (15 μA), with the drain voltage clamped to the positive supply rail, and the source voltage floating with the gate potential. The Ag | AgCl reference electrode, representing the potential in which the floating gate was referred to, was connected to ground. The sensor performance, once integrated in the pill [Fig. 5(a)], corresponded to 14.85 bits pH^{-1} which gave a resolution of 0.07 pH per datapoint. The calibrated response from the pH sensor conformed to a linear regression ($f(pH) = -14.85 pH + 588$, $R^2 = 0.98$), although the sensor exhibited a larger responsivity in alkaline solutions. The sensor lifetime of 20 h was limited by the Ag | AgCl reference electrode made from electroplated silver. The pH sensor exhibited a signal drift of -6 mV h^{-1}.

(0.14 pH), of which -2.5 mV h^{-1} was estimated to be due to the dissolution of AgCl from the reference electrode. The temperature sensitivity of the pH-sensor was measured as 16.8 mV°C^{-1}. Changing the pH of the solution at 40 °C from pH 6.8 to pH 2.3 and pH 11.6 demonstrated that the two channels were completely independent of each other and that there was no signal interference from the temperature channel [Fig. 5(b)].

C. Oxygen Sensor Performance

The electrodes were first characterized using the model redox compound FMCA, showing that the oxygen sensor behaved with classic microelectrode characteristics [26]. The reduction potential of water was subsequently measured at -800 mV (Versus the integrated Ag | AgCl) by recording the steady-state current in oxygen-depleted PBS, thereby excluding any interfering species.

In order to calibrate the sensor, a three point calibration was performed (at saturated oxygen, and with oxygen removed by the injection of Na_2SO_3 to a final concentration of 1 M). The steady state signal from the oxygen saturated solution was recorded at a constant working electrode potential of -700 mV (versus Ag | AgCl), which was below the reduction potential for water. This generated a full-scale signal of 65 nA corresponding 8.2 mg O_2 L^{-1}. The injection of Na_2SO_3 into the PBS after 90 s provided the zero point calibration. This fall in the reduction current provided corroborative evidence that dissolved oxygen was being recorded, by returning the signal back to the baseline level once all available oxygen was consumed. A third, intermediate point was generated through the addition of 0.01 M Na_2SO_3. The resulting calibration graph conformed to a linear regression ($f(\text{mg } O_2) = 7.9 \text{ (mg } O_2)$, $R^2 = 0.99$) with $f(\text{mg } O_2)$ expressed in nanoamperes. The sensitivity of the sensor was 7.9 nA mg^{-1} O_2, with the resolution of 0.4 mg L^{-1} limited by noise or background drift. The lifetime of the integrated Ag | AgCl reference electrode, made from thermal evaporated silver, was found to be to 45 h, with an average voltage drift of -1.3 mV h^{-1} due to the dissolution of the AgCl during operation. Both measurements of FMCA and oxygen redox behavior indicated a stable Ag | AgCl reference.

D. Conductivity Sensor Performance

The prototype circuit exhibited a logarithmic performance from 0.05 to 10 mS cm^{-1} which conformed to a first-order regression analysis ($f(\text{mS cm} - 1) = 165 \ln(\text{mS cm}^{-1}) + 850$, $R^2 = 0.99$) with $f(\text{mS cm}^{-1})$ expressed in millivolts. The sensor saturated at conductivities above 10 mS cm^{-1} due to the capacitive effect of the electric double layer, a phenomena commonly observed in conductimetric sensor systems [29].

E. Control Chip

The background noise from the ASIC corresponded to a constant level of 3-mV peak-to-peak, which is equivalent to one least significant bit (LSB) of the ADC. Since the second LSB were required to provide an adequate noise margin, the 10-bit ADC was anticipated to have an effective resolution of 8 bits.

F. Transmission Frequency

Frequency stabilized units were essential to prevent the transmission drifting out of range, particularly if the pill was subject
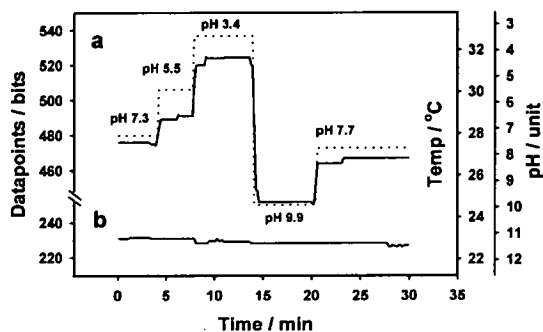
Fig. 6. Recording of pH and temperature *in vitro* using the electronic pill suspended in PBS with the pH and temperature presented on the RHS axis: (a) the solid line represents the acquired data from the pH sensor, with the dotted line representing the real pH as measured using a standard lab pH electrode. An increased signal magnitude corresponds to a reduced pH. The initial pH 7.3 was changed by titrating 0.1 M H_2SO_4 to pH 5.5 (4 min) and pH 3.4 (8 min), respectively. Adding 0.1 M NaOH returned the pH to 9.9 (14 min) before the final pH of 7.7 (20 min) was achieved by titrating 0.1 M H_2SO_4; (b) simultaneous recorded data from the temperature sensor at a constant temperature of 23 °C. The negative drift is due to a reduced supply voltage from the batteries.

to a temperature change during operation. The standard *Type I* transmitter exhibited a negative linear frequency change from 39.17 MHz at 1 °C to 38.98 MHz at 49 °C, corresponding to -4 kHz°C^{-1} ($f(T°_C) = -4 \times 10^{-3}(T°_C) + 39.18$, $R^2 = 0.99$) with the $f(T°_C)$ expressed in Hertz. The narrow signal bandwidth of 10 kHz gave a temperature tolerance of only ± 1.3°C before the signal was lost. In contrast, the *Type II* transmitter exhibited a positive linear frequency change from 20.07 MHz at 2 °C to 20.11 MHz at 45 °C, corresponding to 0.9 kHz°C^{-1} ($f(T°_C) = 9 \times 10^{-4}(T°_C) + 20.07$, $R^2 = 0.99$). Considering the identical signal bandwidth of 10 kHz, the temperature tolerance was increased to ± 5.5°C. The transmitter's signal magnitude was not affected with the pill immersed in the different electrolyte solutions or RO water, compared to the pill surrounded by air only. Tests were also conducted with the pill immersed in the large polypropylene beaker filled with 2000 mL of PBS without the signal quality being compromised. The electromagnetic noise baseline was measured to 78 dB of S/N in the 20 MHz band of the crystal stabilized transmitter.

G. Dual Channel Wireless Signal Transmission

Dual channel wireless signal transmission was recorded from both the pH and temperature channels at 23 °C, with the pill immersed in a PBS solution of changing pH. The calibration graphs for the temperature ($f(T°_C) = 11.9(T°_C) - 42.7$, $R^2 = 0.99$) and pH channel ($f(\text{pH}) = -14.85$ pH $+ 588$, $R^2 = 0.98$) were used to convert the digital units from the MATLAB calculated routine to the corresponding temperature and pH values.

The signal from the pH channel exhibited an initial offset of 0.2 pH above the real value at pH 7.3 [Fig. 6(a)]. In practice, the pH sensor was found to exhibit a positive pH offset as the solution became more acidic, and a negative pH offset as the solution became more alkaline. The response time of the pH sensor was measured to 10 s. The temperature channel was unaffected by the pH change [Fig. 6(b)], confirming the absence of crosstalk between the two channels in Fig. 5(b).

121
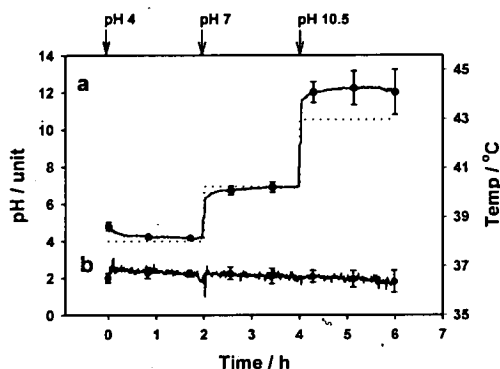
Fig. 7. Long term *in vitro* pH measurements in response to a changing pH from the initial pH 4 to pH 7 (2 h) and pH 10.5 (4 h) at 36.5 °C: (a) solid line represent the recorded data from the pH sensor, with the dotted line representing the real pH as measured using a standard lab pH electrode. The average response illustrates the long-term drift in the sensor after 6 h. The error bars correspond to the standard error of the mean ($n = 2$); (b) the drift from the temperature sensor is solely based on the supply voltage potential, resulting in a smaller error between successive measurements ($n = 2$).

Six hour bench test measurements of the pH and temperature channels in electrolyte solutions, maintained at 36.5 °C, revealed long term drift characteristics of both channels (Fig. 7). The sensor exhibited an initial rapid pH response (30 s), with an additional equilibration time of 2 h required to transmit the correct the pH of the solution. The temperature channel [Fig. 7(b)] exhibited similar drift characteristics as that found in Fig. 6(b).

V. DISCUSSION

All of the components of the sensors and the capsule, exposed to the local environment, had to be able to resist the corrosive environment in the digestive tract, and at the same time be nontoxic (biocompatible) to the organism. If toxic materials were used (such as in batteries and the Ag | AgCl reference electrode), care would need to be taken to prevent leakage from the micro-system and into the surrounding environment.

A. Fabrication

Thermal evaporation of silver generates a dense metal layer, with characteristics closer to bulk metal compared to porous electroplated silver. Although electroplating allow for a thicker layer of silver to be deposited, the lifetime of a Ag | AgCl reference electrode made from 500-nm-thick thermally evaporated silver was compared to a Ag | AgCl electrode made from a 5-μm-thick electroplated layer. The results clearly demonstrated the potential of utilizing thermally evaporated silver in Ag | AgCl electrodes to extend lifetime by more than 100%. However, a protective layer of 20 nm titanium was required to prevent oxidation of the silver in subsequent fabrication levels, and which had to be removed by immersion in a HF solution. Since HF also attacks Si_3N_4, this procedure could not be used in *Chip 1* to avoid damage to the thin 50-nm layer of Si_3N_4 defining the pH sensitive membrane of the ISFET. In contrast, the 500-nm-thick PECVD Si_3N_4 defining the microelectrode array of the oxygen sensor, was tolerant to HF exposure.

The sensor lifetime was further extended through using a three-electrode electrochemical cell for the oxygen sensor, in favor of a two-electrode device. A two electrode unit utilize the reference electrode as a combined counter and reference unit to channel all the current from the reduction of oxygen. However, a three-electrode electrochemical cell bypasses the current flow from the working electrode by incorporating a separate counter electrode subjecting the reference electrode only to the bias current of the input transistor stage of the operational amplifier, to which the sensor is connected. Thus, the overall current channelled through the reference was reduced by at least three orders of magnitude. This effect is important as it enables a reduction in the electrode area and improved long-term stability.

B. Sensor Performance

The temperature circuit was sensitive to the supply voltage. The n-channel of the silicon diode was clamped to ground, whereas the bias current supply circuit was clamped to the negative supply rail. Thus, an increase of 7.25 mV h^{-1} (from a total 14.5 mV h^{-1} from the positive and negative supply rail) would reduce the bias current by 0.5%, resulting in a diode voltage change of -1.6 mV h^{-1} [20]. A potential divider circuit clamped between ground and the positive supply rail was used to create an offset signal prior to the amplification stage. The change in offset signal corresponds to -3.2 mV h^{-1}, resulting in a total signal change of -4.8 mV h^{-1} prior to amplification with a gain of 6.06, resulting in a total change of -29.1 mV h^{-1}. The theoretical calculation conforms to within 40% of the experimental result, which can be explained by real circuit device tolerances (such as supply voltage effect on the operational amplifiers) which deviates from the theoretical predictions.

The pH channel recordings from the pill (Fig. 6) deviated from the true value measured with the glass pH electrode, by transmitting a pH responsivity below the calibrated value. In acidic solutions, this resulted in a pH response slightly above the true value, whereas the response in alkaline solutions was below the true value. In neutral solutions, the pH channel exhibited an offset of 0.2 units above the real value. The results of the long-term measurements conducted in Fig. 7 suggested that the recorded values would match the real pH of the solution if left to equilibrate for 2 h. Thus, the combined effect of calibration offset and short equilibration time to a changing pH, could explain the signal offset between the measured and real pH presented in Fig. 6. The discrepancy between the real and recorded value was possibly due to an inherent memory effect in the pH sensitive Si_3N_4 membrane [30], where the magnitude in response to a changing pH depended on the previous pH value. The difference between the initial pH measurement and the solution value of pH 4 and 7 (Fig. 7) was comparable to the offset magnitudes seen in Fig. 6.

Considering Fig. 7, the offset recorded for pH 10.5 was due to additional factors, such as drift in the reference electrode and supply voltage. The potential divider circuit, which clamped the drain potential of the ISFET was connected between ground and the positive supply rail. Thus, a corresponding change in the positive supply rail of -7.25 mV h^{-1} would result in a drain voltage change of -3.6 mV h^{-1} from the potential divider circuit. The additional drift from the Ag | AgCl reference of -2.5 mV h^{-1} balanced the remaining drift of -6 mV h^{-1}

recorded. The additional discrepancy found at pH 10.5 (Fig. 7) was most likely a result of long-term signal drift from the interaction of proton reactive sites in the bulk of the Si_3N_4 membrane [31], with the drift becoming more predominant in alkaline solutions and at higher temperatures [32].

Bench testing of the oxygen sensor proved satisfactory operation of the electrochemical cell with a low noise (1% of full signal magnitude) and rapid response time of 10 s. However, signal resolution was limited to the standard error of ± 0.4 mg L^{-1}. The signal discrepancy was caused by contamination or deposits on the working electrode surface, which reduced the sensitivity, and by ambient temperature variation, changing the amount of dissolved oxygen by 2% $°C^{-1}$ [33]. Cleaning the surface in an O_2 barrel asher restored the function. However, signal drift was also caused by electrolyte penetration of the interface between the PECVD Si_3N_4 layer and the underlying gold working electrode comprising the microelectrode array. This represented a more serious problem, since it effectively increased the combined surface area of the working electrode resulting in an increase in signal magnitude at a constant dissolved oxygen level.

The conductivity sensor is currently being redesigned to extend the dynamic range. The sensor will be an interdigitated gold planar electrode using Si_3N_4 to prevent the absorption of organic compounds onto its surface.

Methods of digital signal processing will be considered after data acquisition to improve the performance from each sensor with respect to signal drift. In contrast, analog signal algorithms (artificial neural networks) will be used in the sensor electronics to cancel out the memory effect of the pH sensor, and the reduction in sensitivity caused by contamination of the sensor surface.

C. Microsystem

The temperature tolerance of the radio-transmitter excluded dynamic temperature measurements for both the *Type I* and *Type II* pill within the range of the temperature sensor. In order to increase the data collection rate, the potential for signal transmission at the European Industrial, Scientific and Medical (ISM) Standard (433.9 MHz) will be explored [34].

The combined power consumption from the microsystem was higher than the theoretical predictions made during the design process, which considered the implementation of sleep mode and the use of an on-off keying (OOK) transmitter. In the sleep mode, the sensors and analog circuitry were powered up prior to data sampling by the digital processing unit, and then turned off. A constant power mode was used in the experiments, since both the oxygen and pH sensors required time to stabilise (>15 s) after being switched on. Thus, the current consumption from the ASIC and sensors was measured to 54% above the predicted 1.1 mA from the system implementing sleep mode. The FSK type radio transmitters reduced the load from the ASIC by drawing power directly from the batteries rather than the chip, although it consumed on average twice the amount of current than a comparable OOK type transmitter (calculated to 1 mA). The simulated power consumption of the free running *Type I* transmitter (2.45 mA) was 12% above the measured data due to the reduced power consumption at the measured frequency of 39.08 MHz at 25 °C in contrast to the calculated frequency of 40.01 MHz used in the model. The measured current consumption of the crystal

stabilized *Type II* unit was comparable to the modeled data. The capacity of the enlarged SR44 Ag_2O cells used as power supply units could meet the increased current demand, in contrast to the SR26 Ag_2O cells originally proposed.

VI. CONCLUSION

We have developed an integrated sensor array system which has been incorporated in a mobile remote analytical microelectronic pill, designed to perform real-time *in situ* measurements of the GI tract, providing the first *in vitro* wireless transmitted multichannel recordings of analytical parameters. Further work will focus on developing photopatternable gel electrolytes and oxygen and cationselective membranes. The microelectronic pill will be miniaturized for medical and veterinary applications by incorporating the transmitter on silicon and reducing power consumption by improving the data compression algorithm and utilizing a programmable standby power mode.

The generic nature of the microelectronic pill makes it adaptable for use in corrosive environments related to environmental and industrial applications, such as the evaluation of water quality, pollution detection, fermentation process control and the inspection of pipelines. The integration of radiation sensors and the application of indirect imaging technologies such as ultrasound and impedance tomography, will improve the detection of tissue abnormalities and radiation treatment associated with cancer and chronic inflammation.

In the future, one objective will be to produce a device, analogous to a micro total analysis system (μTAS) or lab on a chip sensor [35] which is not only capable of collecting and processing data, but which can transmit it from a remote location. The overall concept will be to produce an array of sensor devices distributed throughout the body or the environment, capable of transmitting high-quality information in real-time.

REFERENCES

[1] S. Mackay and B. Jacobson, "Endoradiosonde," *Nature*, vol. 179, pp. 1239–1240, 1957.
[2] H. S. Wolff, "The radio pill," *New Scientist*, vol. 12, pp. 419–421, 1961.
[3] S. J. Meldrum, B. W. Watson, H. C. Riddle, R. L. Bown, and G. E. Sladen, "pH profile of gut as measured by radiotelemetry capsule," *Br. Med. J.*, vol. 2, pp. 104–106, 1972.
[4] D. F. Evans, G. Pye, R. Bramley, A. G. Clark, T. J. Dyson, and J. D. Hardcastle, "Measurement of gastrointestinal pH profiles in normal ambulant human subjects," *Gut*, vol. 29, no. 8, pp. 1035–1041, Aug. 1988.
[5] R. H. Colson, B. W. Watson, P. D. Fairlclough, J. A. Walker-Smith, C. A. Campell, D. Bellamy, and S. M. Hinsull, "An accurate, long-term, pH sensitive radio pill for ingestion and implantation," *Biotelem. Pat. Mon.*, vol. 8, no. 4, pp. 213–227, 1981.
[6] S. S. Kadirkamanathan, E. Yazaki, D. F. Evans, C. C. Hepworth, F. Gong, and C. P. Swain, "An ambulant porcine model of acid reflux used to evaluate endoscopic gastroplasty," *Gut*, vol. 44, no. 6, pp. 782–788, June 1999.
[7] A. G. Press, I. A. Hauptmann, L. Hauptmann, B. Fuchs, K. Ewe, and G. Ramadori, "Gastrointestinal pH profiles in patients with inflammatory bowel disease," *Aliment Pharm. Therap.*, vol. 12, no. 7, pp. 673–678, Jul. 1998.

[8] G. Pye, D. F. Evans, S. Ledingham, and J. D. Hardcastle, "Gastrointestinal intraluminal pH in normal subjects and those with colorectal adenoma or carcinoma," *Gut*, vol. 31, no. 12, pp. 1355–1357, Dec. 1990.

[9] G. Iddan, G. Meron, A. Glukhovsky, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, no. 6785, p. 417, May 2000.

[10] G. X. Zhou, "Swallowable or implantable body temperature telemeter—body temperature radio pill," in *Proc. IEEE 15th Ann. Northeast Bioeng. Conf.*, Boston, MA, 1989, pp. 165–166.

[11] S. Mackay, "Radio telemetering from within the body," *Science*, vol. 134, pp. 1196–1202, 1961.

[12] E. A. Johannessen, J. M. R. Weaver, L. Bourova, P. Svoboda, P. H. Cobbold, and J. M. Cooper, "Micromachined nanocalorimetric sensor for ultra-low-volume cell-based assays," *Anal. Chem.*, vol. 74, no. 9, pp. 2190–2197, May 2002.

[13] J. W. Gardner, *Microsensors: Principles and Applications*. Chichester, U.K.: Wiley, 1994, pp. 1–331.

[14] W. Gumbrecht, D. Peters, W. Schelter, W. Erhard, J. Henke, J. Steil, and U. Sykora, "Integrated pO2, pCO2, pH sensor system for online blood monitoring," *Sens. Actuators B*, vol. 18–19, no. 1–3, pp. 704–708, Apr. 1994.

[15] C. Belmont-Herbert, M. L. Tercier, J. Buffle, G. C. Fiaccabrino, N. F. de Rooij, and M. Koudelka-Hep, "Gel-integrated microelectrode arrays for direct voltammetric measurements of heavy metals in natural waters and other complex media," *Anal. Chem.*, vol. 70, no. 14, pp. 2949–2956, July 1998.

[16] G. Jobst, G. Urban, A. Jachimowicz, F. Kohl, and O. Tilado, "Thin film Clark-type oxygen sensor based on novel polymer membrane systems for *in vivo* and biosensor applications," *Biosens. Bioelectron.*, vol. 8, no. 3–4, pp. 123–128, 1993.

[17] P. Bergveld, "Development, operation, and application of the ion-sensitive field effect transistor as a tool for electrophysiology," *IEEE Trans. Biomed. Eng.*, vol. BME-19, pp. 342–351, 1972.

[18] G. Asada, A. Burstein, D. Chong, M. Dong, M. Fielding, E. Kruglick, J. Ho, F. Lin, T. H. Lin, H. Marcy, R. MU.K.ai, P. Nelson, F. Newbery, K. S. J. Pister, G. Pottie, H. Sarcelize, O. M. Stafsudd, S. Valoff, G. Young, and W. J. Kaiser, "Low power wireless communication and signal processing circuits for distributed microsystems," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, Hong Kong, 1997, pp. 2817–2820.

[19] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, pp. 114–117, 1965.

[20] M. H. Rashid, *Microelectronics Circuit Analysis and Design*. Boston, MA: PWS, 1999, pp. 1–990.

[21] C. Cane, I. Gracia, and A. Merlos, "Microtechnologies for pH ISFET chemical sensors," *Microelectr. J.*, vol. 28, no. 4, pp. 389–405, May 1997.

[22] L. C. Clark Jr. and C. Lyons, "Electrode systems for continuos monitoring in cardiovascular surgery," *Ann. NY Acad. Sci.*, vol. 102, pp. 29–45, 1962.

[23] L. E. Bermudez, M. Petrofsky, and J. Goodman, "Exposure to low oxygen tension and increased osmolarity enhance the ability of *Myobacterium avium* to enter interstestinal epithelial (HT-29) cells," *Infect. Immun.*, vol. 65, no. 9, pp. 3768–3773, Sep. 1997.

[24] T. Matsuo and K. D. Wise, "An integrated field effect electrode for biopotential recording," *IEEE Trans. Biomed. Eng.*, vol. BME-21, pp. 485–487, 1974.

[25] H. SuzU.K.i, N. Kojima, A. Sugama, F. Takei, and K. Ikegami, "Disposable oxygen electrodes fabricated by semiconductor techniques and their applications to biosensors," *Sens. Actuators B*, vol. 1, no. 1–6, pp. 528–532, Jan. 1990.

[26] M. E. Sandison, N. Anicet, A. Glidle, and J. M. Cooper, "Optimization of the geometry and porosity of microelectrode arrays for sensor design," *Anal. Chem.*, vol. 74, no. 22, pp. 5717–5725, Oct. 2002.

[27] L. Wang, T. B. Tang, E. A. Johannessen, A. Astaras, A. F. Murray, J. M. Cooper, S. P. Beaumont, and D. R. S. Cumming, "An integrated sensor microsystem for industrial and biomedical applications," in *Proc. IEEE Instrumentation Measurement Technology Conf.*, vol. 2, Anchorage, AK, 2002, pp. 1717–1720.

[28] R. M. C. Dawson, D. C. Elliott, W. H. Elliott, and K. M. Jones, *Data for Biochemical Research*, 3rd ed. Oxford, U.K.: Clarendon, 1986, pp. 1–580.

[29] H. Morgan and N. Green, *AC Electrokinesis: Colloids and Nanoparticles*. Baldock, U.K.: Research Studies Press Ltd, 2003, pp. 1–324.

[30] L. Bousse, D. Hafeman, and N. Tran, "Time-dependence of the chemical response of silicon nitride surfaces," *Sens. Actuators B*, vol. 1, no. 1–6, pp. 361–367, Jan. 1990.

[31] D. Yu, Y. D. Wei, and G. H. Wang, "Time-dependent response characteristics of pH-sensitive ISFETS," *Sens. Actuators B*, vol. 3, no. 4, pp. 279–285, Apr. 1991.

[32] P. Hein and P. Egger, "Drift behavior of ISFET's with $Si_3N_4 - SiO_2$ gate insulator," *Sens. Actuators B*, vol. 13–14, no. 1–3, pp. 655–656, June 1993.

[33] M. Hitchman, *Measurement of Dissolved Oxygen*. New York: Wiley, 1978, pp. 1–255.

[34] L. C. Chirwa, P. A. Hammond, S. Roy, and D. R. S. Cumming, "Electromagnetic radiation from ingested sources in the human intestine between 150 MHz and 1.2 GHz," *IEEE Trans. Biomed. Eng.*, vol. 50, pp. 484–492, Apr. 2003.

[35] P. Wilding and L. J. Kricka, "Micro-microchips: just how small can we go?," *Trends Biotechnol.*, vol. 17, no. 12, pp. 465–468, Dec. 1999.

Erik A. Johannessen received the B.Sc. degree in natural sciences from the University of Tromsø, Tromsø, Norway, in 1997, and the Ph.D. degree from the University of Liverpool, Liverpool, U.K., in 2002. The PhD project (Wellcome Trust) focused on ultra-small nanocalorimetric sensors for high-density assay screening of cells.

He is currently a Postdoctoral Research Associate with the Department of Electronic and Electrical Engineering, University of Glasgow, Glasgow, U.K. His research interests is in bioelectronics and nanoelectronics, silicon microfabrication, and analytical microsystem design.

Lei Wang (M'03) received the B.Sc. degree in information and control engineering and the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 1995 and 2000, respectively.

He was an academic visitor in the Department of Mechanical Engineering at the University of Dundee, U.K., and is currently a Postdoctoral Research Associate with the Department of Electronic and Electrical Engineering, University of Glasgow, Glasgow, U.K.. His research interests is in physiological measurement systems, digital signal processing, and integrated circuit design.

Li Cui received the undergraduate degree from Tsinghua University, China, in 1985 and the M.Sc. degree from the Institute of Semiconductors, Chinese Academy of Sciences, in 1988, where she worked on solid state and electrochemical sensors. She received the Ph.D. degree in electronic nose applications from the University of Glasgow, Glasgow, U.K., 1999.

She is currently a Postdoctoral Research Associate with the Department of Electronic and Electrical Engineering, University of Glasgow, Glasgow, U.K. Her research interests is in biosensors, dielectrophoresis, and packaging for lab-on-a-chip applications.

Tong Boon Tang (S'02) received the B.Eng. (Hons) degree in electronics and electrical engineering communications from the University of Edinburgh, Edinburgh, U.K., in 1999. He is working towards the Ph.D. degree in intelligent sensor fusion part time at the University of Edinburgh.

He was ASIC engineer with the Singapore Design Centre, Lucent Technologies, and joined the Integrated Systems Group at The University of Edinburgh as a Research Assistant in 2001. His research interests lie in unsupervised stochastic neural algorithms, analog VLSI design, and smart sensors.

Mansour Ahmadian received the B.Sc. degree in electronic engineering from the K.N.Toosi University, Tehrah, Iran, in 1989, and the M.Sc. degree from Sharif University, Tehran, in 1992. He received the Ph.D. in biomedical engineering from the University of Edinburgh, Edinburgh, U.K., in 2002.

He is currently a Postdoctoral Research Assistant with the the School of Engineering and Electronics, University of Edinburgh. His research interests are medical signal and image processing, parallel processing, communication systems, and microwave design.

Alexander Astaras (S'02) received the B.A. degree in physics from Oberlin College, Oberlin, OH, in 1995. He is currently pursuing the Ph.D. degree in analog VLSI circuit design for unsupervised probabilistic neural network part-time in the School of Engineering and Electronics at the University of Edinburgh, Edinburgh, U.K.

He is a Research Assistant with the Institute for System Level Integration in Livingston, U.K., on system-on-chip design. His research interests are diagnostics, instrumentation, neural networks, and mixed-signal system-on-chip VLSI design.

Stuart W. J. Reid received the BVetMed and Ph.D. degrees from the University of Glasgow, Glasgow, U.K., in 1987 and 1992, respectively.

He is a Diplomate at the European College of Veterinary Public Health and a Fellow of the Royal Society of Edinburgh. He is currently a Professor of Comparative Epidemiology and Informatics with the Univeristies of Glasgow and Strathclyde (Strathclyde, U.K.). His current research interests comprise epidemiology, informatics, and statistical modeling of disease and disease processes.

Philippa S. Yam received the B.Sc. (Hons) degree in veterinary science in 1990 and the BVM&S degree from the University of Edinburgh, Edinburgh, U.K., in 1992. She received the Ph.D. degree from the University of Glasgow, Glasgow, U.K., in 1999.

She is a Hill's Lecturer in Gastroenterology at the University of Glasgow Veterinary School. Her research interests focus on gut motility and the development of noninvasive diagnostics.

Alan F. Murray (M'91–SM'93) received the B.Sc. (Hons) degree in physics and the Ph.D. degree in solid state physics from the University of Edinburgh, Edinburgh, U.K.,in 1975 and 1978, respectively.

He has been a Professor of of Neural Electronics in the School of Engineering and Electronics at the University of Edinburgh since 1994. His research interests comprise (probabilistic) neural computation, hardware-compatible learning schemes, neural network applications, and the interface between silicon and neurobiology.

Dr. Murray is a Fellow of the Institution of Electrical Engineers (IEE) and the Royal Society of Edinburgh.

Brian W. Flynn received the B.Sc. and Ph.D degrees in electrical engineering from the University of Edinburgh, Edinburgh, U.K., in 1971 and 1977, respectively.

He is a Senior Lecturer in the Department of Electrical Engineering at the University of Edinburgh, and a Development Engineer with Marconi Communications Systems Ltd., Livingston, U.K., specializing in microwave radio link systems. He is the co-author of a textbook on the design of switched mode power supplies. His current research interests comprise electromagnetics and RF/microwave design.
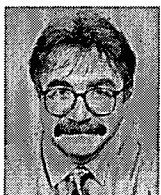
Steve P. Beaumont received the M.A. and Ph.D. degrees in electrical sciences from the University of Cambridge, Cambridge, U.K., in 1972 and 1979, respectively.

He is presently Director of the Institute for System Level Integration, which he founded in 1998, and Professor of Nanoelectronics at the University of Glasgow, Glasgow, U.K. His research interests have covered silicon nanofabrication, millimeter-wave integrated circuit design, and systems-on-chip technology including integration with peripheral components such as sensors and lab-on-chip devices.
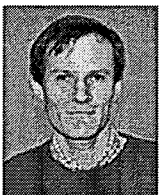
David R. S. Cumming (M'97) received the B.Eng. degree in electronic engineering from the University of Glasgow, Glasgow, U.K., 1989 and the Ph.D. degree from the University of Cambridge, Cambridge, U.K., in 1993.

He has worked on mesoscopic device physics, RF characterization, diffractive optics for optical and submillimeter wave applications, diagnostic systems, and microelectronic design. He is a Senior Lecturer and EPSRC Advanced Research Fellow in Electronics and Electrical Engineering at the University of Glasgow, where he leads the Microsystem Technology Group.

Jonathan M. Cooper received the B.Sc. degree in biological sciences from the University of Southampton, Southampton, U.K., in 1983 and Ph.D. degree in sensor technology from the University of Cranfield, Cranfield, U.K., in 1989.

He is a Professor of Bioelectronics in the Department of Electronic and Electrical Engineering at the University of Glasgow, Glasgow, U.K. His research interests is in medical diagnostics, thermal, electrochemical and biochemical detection methods, and lab-on-a-chip devices.

Prof. Cooper is a Fellow of the Royal Society of Edinburgh the Institute of Physics and the Institution of Electrical Engineers (IEE).

125

Adaptive, integrated sensor processing to compensate for drift and uncertainty: a stochastic 'neural' approach

T.B. Tang, H. Chen and A.F. Murray

Abstract: An adaptive stochastic classifier based on a simple, novel neural architecture - the Continuous Restricted Boltzmann Machine (CRBM) is demonstrated. Together with sensors and signal conditioning circuits, the classifier is capable of measuring and classifying (with high accuracy) the H^+ ion concentration, in the presence of both random noise and sensor drift. Training on-line, the stochastic classifier is able to overcome significant drift of real incomplete sensor data dynamically. As analogue hardware, this signal-level sensor fusion scheme is therefore suitable for real-time analysis in a miniaturised multisensor microsystem such as a Lab-in-a-Pill (LIAP).

1 Introduction

Rapid progress in both Lab-on-a-Chip (LOC) and System-on-Chip (SoC) technologies has encouraged increasing interest in electronic health care [1]. Applications have ranged from telemedicine to bioanalysis, from patient monitoring to implantable devices. In conjunction with image scanning [2], several other biomedical instruments provide constant monitoring of such essential physiological parameters as temperature, pH, oxygen and pressure [3–6].

It is clearly desirable that such instruments be integrated and thus miniaturised, although it is inherently more difficult to extract useful information from what are now far more noisy and unstable measurements. With integration, however, comes the possibility of sensor redundancy (multiple sensors of the same or different types). There is, therefore, a need for robust, adaptive algorithms for sensor fusion and early pre-processing, that can be implemented directly in hardware, with low power consumption. In particular, algorithms that can process continuous-time, analogue sensor signals directly are especially useful [7].

This paper investigates the ability of stochastic neural computation to fuse multisensor data at signal level under conditions of significant sensor drift. Figure 1 shows two separate sets of measured, and typical, sensor drift (over 21 h) in 10 pH-ISFET sensors. The data is obtained by measuring the potential of several drifting reference electrodes in a neutral (pH 7) buffer over time. The drift, which causes the drop in reference potential, is due to the dissolution of AgCl from the reference electrode [8]. In real applications, this drift will destroy the capabilities of the sensor system unless regular recalibration, or some form of adaptive self-calibration, is introduced. We are interested in the latter approach. An adaptive classifier must be able both to track sensor drift and to maintain its classification ability

in the absence of a complete, representative training set (i.e. only data drawn from a sub-class of the full data space are likely to be available at one time, under normal operating conditions). This is a serious challenge. Without constrained
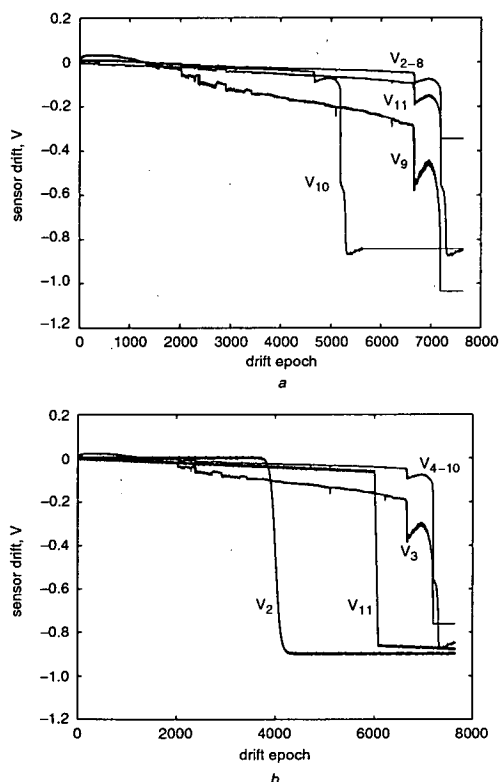
Fig. 1 *Two separate sets of sensor drift in several reference electrodes over time*

28

training, most continuously-adaptive systems will simply "learn" to model the current distribution, losing the ability to model the entire data space completely. Such systems thus lose classification ability abruptly and completely.

In this paper, we describe the adaptive stochastic classifier, and discuss training and some important practicalities in this application. Experimental results that demonstrate the feasibility of this approach and compare its performance with both linear and nonlinear (multi-layer perceptron) classifiers are presented. Both are also "neural" architectures that can be trained to classify data, but are not continuously-adaptive [9].

2 Architecture

The classifier's operation can be viewed in two stages; unsupervised feature extraction and supervised (linear) classification. Our aim is to render the feature-extraction stage adaptive and able to present a consistent set of features to the supervised classifier in the presence of drift and noise. The Continuous Restricted Boltzmann Machine (CRBM) [10] is a generative model that is capable of a form of autonomous feature extraction and is based upon Hinton's product-of-experts architecture [11]. The CRBM has one visible layer that passes and receives data to and from the outside world and one hidden layer that underpins the ability to model the mechanisms that underlie a set of data (Fig. 2). This algorithm is specifically designed to accept analogue continuous signals and to build from them a continuous-valued generative model.

In the absence of a full physical/chemical/biological model, a generative model is a useful statistical (often "neural") model that is trained to generate, or to "model" data with the same statistical structure as a set of "training data". A good generative model mimics both the structure and inherent noisiness of real, multi-dimensional data. It is usable for classification and novelty detection of unseen data drawn from the same physical source as the training data, as the model can be viewed as providing an "explanation" of the mechanism(s) that generated the data.

The CRBM thus avoids quantisation and the loss of information in, for example, the binary Restricted Boltz-mann Machine (RBM) [11, 12]. Furthermore, the CRBM's simple computation requires only local addition and multiplication, and is thus (analogue) hardware amenable [10]. The CRBM is trained in an unsupervised manner, adapting both its weights (internal model parameters) and internal noise sources by minimising contrastive divergence [10, 11].

The final, output, block of the classifier used in this paper is a single layer perceptron (SLP). In our application, there is one temperature sensor [13] and 10 pH-ISFET sensors [14] which measure the environmental concentration of H^+ ions. The temperature sensor monitors the ambient temperature, to improve overall system robustness, as the pH sensors are also temperature-dependent. The use of redundant pH-ISFET sensors aims to improve the overall robustness of the classifier. Each sensor output is passed directly to a visible unit i in the CRBM. The 11 visible units are connected to and from the neurons j in the hidden layer via a symmetrical weight matrix w_{ij}. In addition, two permanently-on bias, or "threshold" units V_0 and H_0 encode, in the weights that lead from them, the adaptive thresholds, or biases, of the hidden and visible units respectively. It will be seen that much of the modelling ability is found in these weights. In the hidden layer, a total of 4 hidden units/neurons aim to encode the mechanisms behind the sensor data distribution. Smaller numbers of hidden units result in poorer models, with too few degrees of freedom. Although more hidden units offer the promise of a better model, we wish to minimise computation and therefore network size. Furthermore, Occam's Razor suggests that a parsimonious model is most likely to be useful and informative [15]. The activity for each stochastic neuron is given by:

$$s_j = tanh \left(a_j \cdot \left(\sum_i w_{ij} s_i + \sigma \cdot N_j(0, 1) \right) \right) \qquad (1)$$

where

$s_i =$ input from neuron i

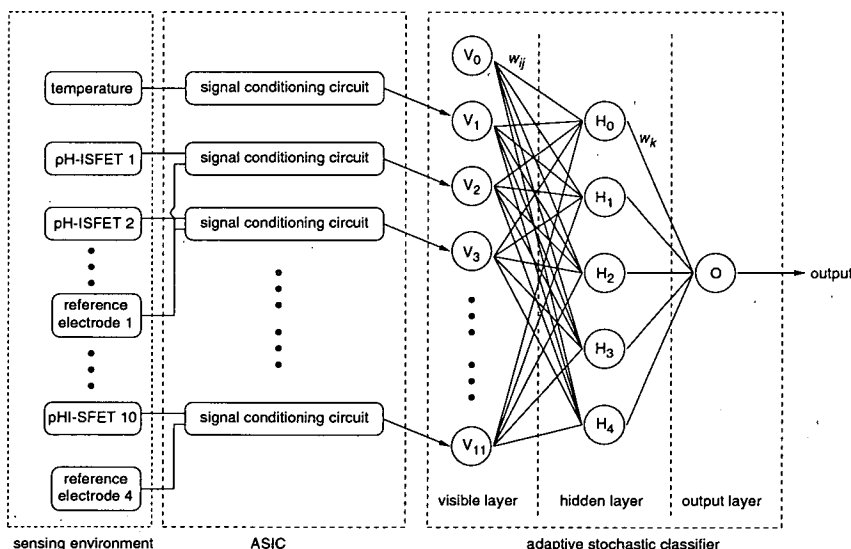$a_j =$ noise control parameter: specific to each (visible or hidden) unit

Fig. 2 An adaptive stochastic classifier in a typical multisensor microsystem with 11 visible, 4 hidden, 2 bias and 1 output units

IEE Proc.-Nanobiotechnol. Vol. 151, No. 1, February 2004

29

127

σ = noise scaling constant: specific to each (visible or hidden) layer

$N_j(0, 1)$ = sampled from unit-magnitude, zero-mean Gaussian noise source

There are three sets of learning parameters in the classifier. They are the weights w_{ij}, the noise control parameters, a_j of the CRBM and the weights w_k in the output SLP layer. The CRBM parameters are optimised by minimising contrastive divergence (MCD) [11] while the SLP is trained using the delta rule [16].

3 Methodology

The signal conditioning circuits can be tuned to maximise sensor sensitivity and thus the separation of any clusters in the sensor data. This facilitates data modelling. However, variance in the threshold voltage of the pH-ISFET sensors due to the fabrication process means that very few integrated sensors of this type can respond linearly across the full pH range. A compromise is therefore made between sensor sensitivity and linearity that renders these integrated sensors slightly nonlinear. The CRBM is fundamentally a nonlinear model and is able to deal with at least this level on nonlinearity, as will be demonstrated. Based on measurements [17], a particular pH-ISFET sensor, when immersed in a solution with pH x, can be modelled as an output voltage $y = A - Bx$ (mv) with a correlation coefficient[1] $R^2 = 0.98$. Typically, $A = 1733.66$ and $B = 43.51$. Within an array of pH-ISFET sensors, threshold voltages vary, the constant term A varies, while the sensitivity B remains fairly consistent. The temperature sensor can be modelled with an output voltage of $y = 34.86T - 125.10$ (mv) for temperature T with $R^2 = 0.99$.

Figure 3 shows the training process for the adaptive stochastic classifier. Initially, the CRBM is trained with 2 datasets. Dataset A comprises measurements on a solution with a temperature of 37°C and pH 4. Dataset B consists of
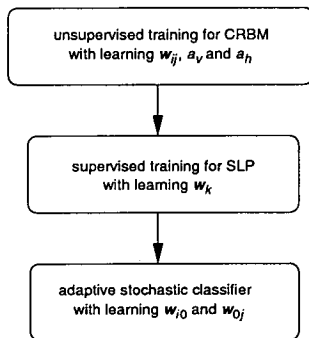

Fig. 3 *Sequential training for the adaptive stochastic classifier*

measurements on a solution with a temperature of 37°C and pH 10. To facilitate training, the noise scaling constant σ is set to a value that avoids both the over-fitting that is associated with low noise and the complete loss of modelling ability that is produced by high noise. Empirical experiments lead to optimal values of σ of 0.2 for visible

units and 0.4 for hidden units. These optimal values are problem-dependent.

The learning rate for the CRBM must also be considered carefully. Our empirical "rule of thumb", drawn from several CRBM-modelling projects, is to have the visible noise control parameters' learning rate η_v 10 times that of hidden units and weights, η_h and η_w, respectively. This encourages autonomous annealing through adaptation of the visible layer's noise control parameters a_v on a shorter timscale than that for adaptation of the weights w_{ij} and hidden noise control parameters a_h to model the detail of training data distribution.

"Greedy training" is used to train the CRBM and SLP layers. CRBM training is allowed to reached equilibrium and is then stopped. At this stage, the SLP is untrained. The SLP is then trained to map the activity of the CRBM hidden layer (H_{0-4}), as the input data is presented to the visible layer, to the known-correct SLP output classification. This training is performed with (a) the visible units clamped to datasets A and B in the training set, (b) the output units clamped to the corresponding labels and (c) fixed weight w_{ij}, noise control parameters a_v and a_h. The only learning parameters are the weights w_k for the output unit.

After training, the learning rates for all parameters except the CRBM bias weights are set to zero. This setting allows the classifier to adapt to sensor drift via w_{i0}^2 and suppresses the competitive learning in CRBM that would otherwise generate a totally new distribution (in this case, one cluster instead of two) and hence destroy classification. Secondly, it causes the representations (the activities of hidden units) that are the extracted "features" passed to the subsequent layer to be consistent for a particular dataset in the presence of drift. Equation (1) shows that the hidden unit's state s_j drifts as the sensor output s_i drifts. Therefore, the visible bias unit's weight w_{0j}, which encodes the biases for all the hidden units, must be allowed to adapt to compensate. Δw_{i0} accounts for a shift in the mean of data distribution, so

$$w_{0j}(t) = w_{0j}^* - \eta_w \sum_i (w_{ij}^* \{w_{i0}^* - w_{i0}(t)\}) \qquad (2)$$

where $i \neq 0$ and $j \neq 0$. w^* refers to weight after the CRBM is trained and before it is exposed to drifting data. Hence the weight change for a visible bias unit at time $t+1$ is:

$$\Delta w_{0j}(t+1) = \eta_w \sum_i (w_{ij}^* \{w_{i0}(t+1) - w_{i0}(t)\}) \qquad (3)$$

where $i \neq 0$ and $j \neq 0$.

4 Experimental results and discussion

The classifier has been trained with two datasets (each with 400 samples) using real sensor data. The CRBM and the SLP are trained for 3000 and 500 epochs respectively. Training results are discussed in section 4.1, while section 4.2 discusses the classifier's ability to track sensor drift via constrained on-line adaptation of the CRBM. The experiment in section 4.2 is conducted by introducing a typical pattern of drift to the dataset A and presenting the data to the visible units of the CRBM. The data is taken over a period of 76440 s, sampled at 0.1 Hz. There are therefore 7644 "drift epochs" in the experiment. Note that no samples from dataset B are presented to the classifier for the experiment in section 4.2.

---

[1]In this context, correlation coefficient is a quantity which gives the quality of a least squares fitting to the measurements.

[2]Experiment [10] has shown that the hidden bias unit's weights acts as an encoder for the mean of the training data distribution. This is due to its state (permanently '+1') which allows it to learn faster (with a larger weight change Δw_{i0}) than other hidden units that have near-zero initial states.

30

IEE Proc.-Nanobiotechnol. Vol. 151, No. 1, February 2004

128

4.1 Learning to classify

Figure 4a and b show 2-dimensional plots of sensor signals for V_1 (temperature sensor), V_2 (pH-ISFET sensor 1) and V_3 (pH-ISFET sensor 2). Sensor noise causes significant overlap between the two clusters in all 11 dimensions. Figure 4c shows the evolution of the CRBM's parameters a_v which suggest that training equilibrium is reached within 3000 training epochs.

It is a characteristic of the CRBM that the stochastic hidden units can adapt to become more or less binary [10]. After training, hidden unit H_2, in this particular case, has a large noise control parameter ($a_h = 2.1$) and thus behaves as a binary "decision" unit that captures gross structure in the data while other hidden units' a_h remain approximately 1, rendering their behaviour more deterministic. More deterministic units are able to model finer detail in the data distribution. Similarly, the output unit's weight w_k connecting to the hidden unit H_2 has increased significantly to 6.2 after 500 training epochs, as shown in Fig. 4e.

These parameters hold the clue as to how the CRBM models these two 11-dimensional clusters of data. The large values imply that the activities of hidden unit H_2 and output unit are very sensitive to the particular elements in the sensor data space. The trained CRBM is then tested with two new datasets (400 samples for each sub-class). As indicated in Fig. 4f, the clear separation between the output response for the two datasets yields 100% accurate classification by simply thresholding at zero. To investigate the classifier's performance further, the above experiment has been repeated using solutions with different pH values in pairs. Each pair has one acidic and one alkaline solution. Figure 5a shows the corresponding output unit's response. In the worst case (i.e. small separation between data clusters),

the first solution (dataset A) has a pH of 6 while the second (dataset B) has pH 8. A 2-dimensional plot of sensor outputs for visible units V_1 and V_2 is shown in Fig. 5b. Despite overlap between the data clusters, the CRBM/SLP still classifies with 81.75% accuracy, as shown in Fig. 5c and d, as the CRBM's ability to model the shape and spread of a data cluster captures structure and "shapes" in the data that a simple linear or nonlinear classifier alone does not.

4.2 Tracking sensor drift

As shown in Fig. 1a, there are three interesting phases in this temporal data stream. The first phase ends at the 4000th drift epoch where the classifier must adapt to gradual sensor drift for each sensor output. The second phase ends at the 6000th drift epoch when the ninth pH-ISFET sensor (V_{10}) fails catastrophically. Finally, all remaining pH-ISFET sensors fail around the 7644th drift epoch.

Figure 6a shows the evolution of w_{i0} which "follows" the sensor drift. The CRBM section of the classifier is compensating autonomously for the drift. However, around the 7644th drift epoch, there is a major shift in several sensor outputs and a failure in classification, indicating that the CRBM model has broken down completely and is now, quite correctly, responding to the input signal as completely new data. This is obviously caused by simultaneous failure in all remaining pH-ISFET sensors.

Figure 6b shows the activity of the output unit as sensor drift occurs and the CRBM's constrained adaptation responds to the gradual change in sensor activity. Ideally, a straight line (output = 1) should be obtained, as only dataset A is presented. However, any sudden and significant increase in the speed of drift will result in a temporary loss of accurate classification as the CRBM adapts to take
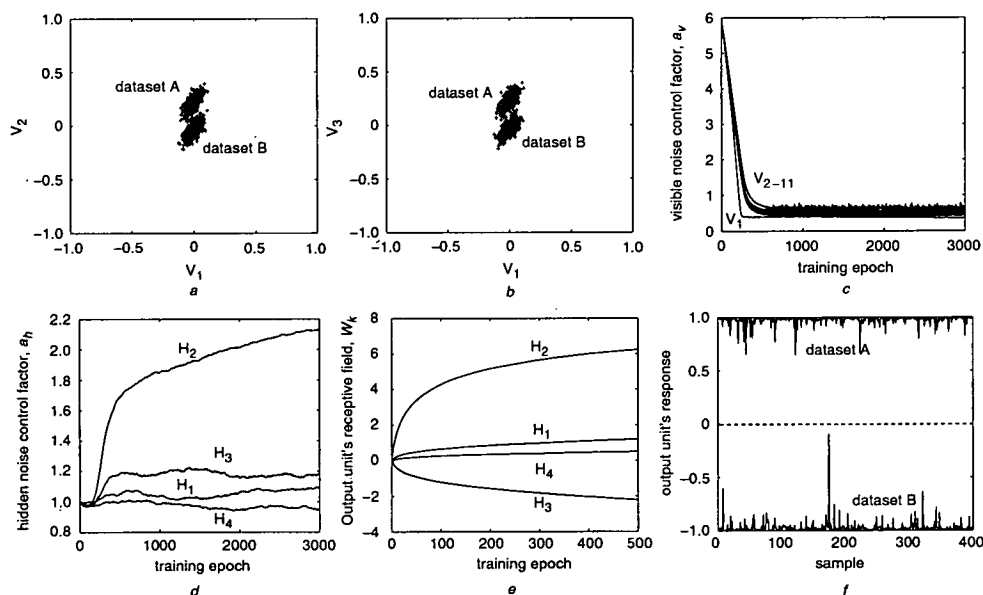
Fig. 4 *Learning in the classifier*
a The training data for visible units V_1 and V_2
b The training data for V_1 and V_3
c The visible noise control parameter a_v for CRBM over 3000 training epochs
d The hidden noise control parameter a_h for CRBM over 3000 training epochs
e The weights w_k for the output unit in SLP over 500 training epochs
f The output unit's response with respect to datasets A and B. The learning rates are 0.1 for weight vector and 3 for noise control parameters

IEE Proc.-Nanobiotechnol. Vol. 151, No. 1, February 2004

31
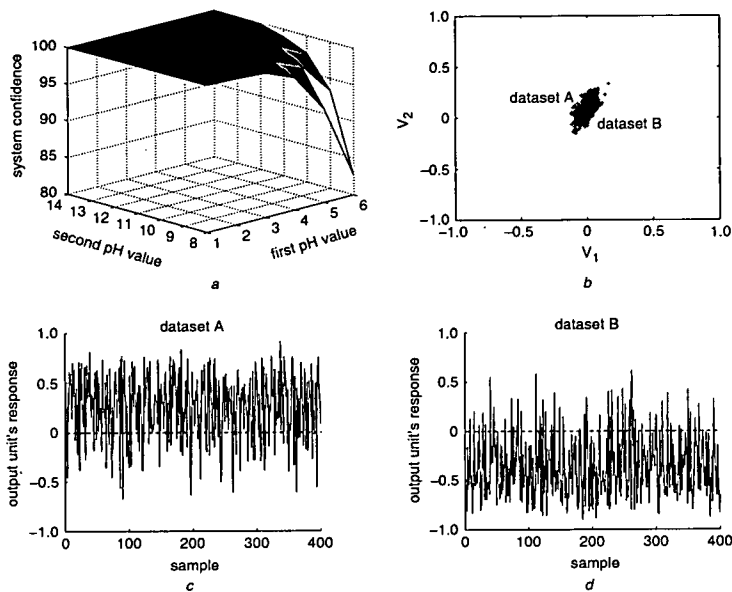
Fig. 5 *Classifier performance over resolution*
a The compiled result for various pairs of solutions
b The training data for visible units V_1 and V_2 with first pH = 6 (dataset A) and second pH = 8 (dataset B)
c The output unit's response for dataset A in the worst case
d The output unit's response for dataset B in the worst case
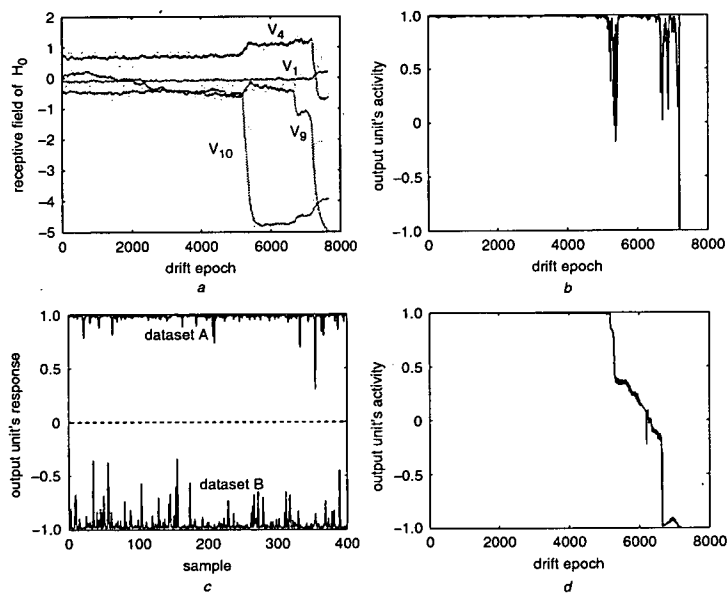
Fig. 6 *Evolution of w_{i0}, activity of output unit as sensor drift occurs, classifier response at 6000th drift epoch and MLP classifier response during 7644 drift epochs*
a Weight change for hidden bias unit w_{i0}
b The output unit's activity during the 7644 drift epochs
c The classifier's response to the two datasets at 6000th drift epoch
d MLP classifier's output response during the 7644 drift epochs.
Experiment is carried out based on drift data shown in Fig. 1*a*

account of the speedier change. This is due to the use of a constant learning rate for the CRBM- a common problem for on-line learning in a dynamic environment [18–20]. Otherwise, the classifier is working as intended and well.

Figure 6*c* shows its response to two datasets (400 new samples for each sub-class) at the 6000th drift epoch. 100% classification accuracy is achieved very simply by thresholding the SLP output at zero.

To highlight the significance of the classifier's on-line unsupervised learning, trained, but subsequently non-adaptive, linear (another SLP) and MLP classifiers are used as benchmarks. The SLP is trained with the two original datasets A and B (i.e. without drift data) for 500 training epochs. It has a learning rate of 0.10 and a sigmoidal activation function. The trained SLP cannot classify even the initial data with 100% accuracy, highlighting the (modest) inherent nonlinearity of the classification task. The classification ability subsequently collapses as sensor drift occurs. The comparison in performance is summarised in Table 1. The MLP has 15 hidden and 1 output units. All units have sigmoidal activation function and use a back-propagation (BP) gradient descent with momentum learning rule [21]. The learning rate is 0.05 and the momentum factor is 0.9. The MLP is trained with the two original datasets for 20000 learning epochs and achieves a mean square error (MSE) of 4.92×10^{-4}, before being

Table 1 Classifiers' output and accuracy at various drift epochs for drift data in Fig. 1a

Parameter	Method	Drift epochs		
		4000th	6000th	7644th
Output	Linear	0.78	−0.18	−1.00
	MLP	0.99	0.15	−1.00
	CRBM+SLP	0.98	0.96	1.00
Accuracy (%)	Linear	99.875	68.000	50.000
	MLP	99.625	77.375	50.000
	CRBM+SLP	99.875	95.500	50.000

exposed to drifting data. The output unit's activity is recorded throughout the drift epochs by Fig. 6d. The MLP's classification ability also collapses during sensor drift, while the CRBM maintains 95.5% correct classification during gentle sensor drift (up to the 6000th epoch). Naturally, all classifiers fail when the sensor activity changes dramatically around 7600th epochs.

The above experiment is repeated with a second set of drift data (Fig. 1b) to confirm that the above result is not an accident. The datastream is divided into four phases. The first ends at the 3500th drift epoch where the classifier is adapting to gradual sensor drift. The second phase ends at the 5000th drift epoch when the first pH-ISFET sensor (V_2) fails. The third phase ends at the 6400th drift epoch when the tenth pH-ISFET sensor (V_{11}) fails. The remaining drift epochs then form the last phase. The experimental results are shown in Fig. 7. As in the previous experiment, the CRBM/SLP classifier compensates for sensor drift as well as sensor failure if sufficient time for adaptive recovery is allowed. If several sensors fail simultaneously or within a very short time (with respect to the learning rate), the CRBM/SLP classifier will fail. Table 2 summarises the results of this second experiment, highlighting once more the CRBM's ability to "track" sensor drift that destroys the performance of other trained, but non-adaptive, classifiers.

5 Conclusions

The implementation of an adaptive classification system has been presented in the context of an 11-dimensional microsystem application. The results show that a CRBM with 4 hidden units is able to track nonlinear sensor drift and to classify noisy sensor data accurately and for far

Fig. 7 *Evolution of w_{i0}, activity of output unit as sensor drift occurs, classifier response at 6400th drift epoch and MLP classifier response during 7644 drift epochs*
a Weight change for hidden bias unit w_{i0} during the 7644 drift epochs
b The output unit's activity during the 7644 drift epochs
c The classifier's response to the two datasets at the 6000th drift epoch
d MLP classifier's output response during the 7644 drift epochs.
Experiment is carried out based on drift data shown in Fig. 1b

IEE Proc.-Nanobiotechnol. Vol. 151, No. 1, February 2004

33

Table 2 Classifiers' output and accuracy at various drift epochs for second set of drift data in Fig. 1b

| Parameter | Method | Drift epochs | | |
		3500th	5000th	6400th
Output	Linear	0.87	0.02	−0.85
	MLP	1.00	0.59	−1.00
	CRBM+SLP	0.99	0.90	0.92
Accuracy (%)	Linear	99.875	76.000	50.750
	MLP	100.000	84.250	50.000
	CRBM+SLP	99.875	96.500	98.500

longer than does a carefully-trained, but non-adaptive neural classifier. Most importantly, the CRBM can be configured to respond sensibly to incomplete and "unbalanced" real-time input data that do not adequately represent the distribution of the training data.

We have also studied the classifier's ability to cope with major faults in the sensors. It has been tested rigorously with noisy real data including sensor outputs that drift at diverse rates and with faulty sensor outputs. Under such circumstances, both linear and MLP non-adaptive classifiers fail immediately. The CRBM classifier and training approach proposed in this paper recovers from major, nonlinear drift within a short period of time. This is particularly useful and important in the chemical sensing application that forms the motivation for this study, where robustness in the face of miniature, noisy and unreliable sensors is the primary challenge. The work in this paper reinforces a more general capability, that a suitable generative model with a constrained training process can adapt to at least some environmental changes such as sensor drift while presenting a consistent (effectively autonomously recalibrated) representation of drifting data to subsequent layer(s) of processing.

Although the proposed classifier has only been applied to clusters that are only modestly nonlinear, its extension to completely nonlinear data distributions is also possible and has been examined in the context of artifical data. This research work is currently still in progress and will be reported in a subsequent paper with real data.

6 Acknowledgments

7 References

1 Aguilo, J., Millan, J., and Villa, R.: 'Micro and nano technologies in medical applications: a challenge'. Proceedings of the International Semiconductor Conference, Sinaia, Romania, 2001, Vol. 1, pp. 247–255
2 Given Imaging Ltd, http://www.givenimaging.com, 2001
3 Zhou, G.X.: 'Swallowable or implantable body temperature telemeter - body temperature radio pill'. Proceedings of 15th Annual Northeast Bioengineering Conference, Boston, MA, USA, 1989, pp. 165–166
4 Evans, D.F., Pye, G., Bramley, R., Clark, A.G., Dyson, T.J., and Hardcastle, J.D.: 'Measurement of gastrointestinal pH profiles in normal ambulant human subjects', Gut, 1988, 29, pp. 1035–1041
5 Jobst, G., Urban, G., Jachimowicz, A., Kohl, F., Tilado, O., Lettenbichler, I., and Nauer, G.: 'Thin film Clark-type oxygen sensor based on novel polymer membrane systems for in vivo and biosensor applications', Biosens. Bioelectron., 1993, 8, pp. 123–128
6 Mackay, S.: 'Radio telemetering from within the body', Science, 1961, 134, pp. 1196–1202
7 Murray, A.F., and Woodburn, R.J.: 'The prospects for analogue neural VLSI', Int. J. Neural Syst., 1998, 8, (5), pp. 559–580
8 Johannessen, E.A., Wang, L., Cui, L., Tang, T.B., Ahmadian, M., Astaras, A., Reid, S.W., Yam, P., Murray, A.F., Flynn, B.W., Beaumont, S.P., Cumming, D.R.S., and Cooper, J.M.: 'Implementation of distributed sensors in a microsystems format', IEEE Trans. Biomedical Eng., 2003, (in press)
9 Ackley, D.H., Hinton, G.E., and Sejnowski, T.J.: 'A learning algorithm for Boltzmann machine', Cogn. Sci., 1985, 9, pp. 147–169
10 Chen, H., and Murray, A.F.: 'A continuous restricted Boltzmann machine with a hardware-amenable learning algorithm'. Proceedings for the 12th International Conference on Artificial Neural Networks, Madrid, Spain, 2002, pp. 358–363
11 Hinton, G.E.: 'Training products of experts by minimizing contrastive divergence', Neural Comput., 2002, 14, (8), pp. 1771–1800
12 Murray, A.F.: 'Novelty detection using products of simple experts - a potential architecture for embedded systems', Neural Netw., 2001, 14, pp. 1257–1264
13 Rashid, M.H.: 'Microelectronics circuit analysis and design' (PWS Publishing Company, Boston, MA, USA, 1999)
14 Bergveld, P.: 'Development, operation and application of the ion-sensitive field effect transistor as a tool for electrophysiology', IEEE Trans. Biomed. Eng., 1972, 19, pp. 342–351
15 Rasmussen, C.E., and Ghahramani, Z.: 'Occam's razor', Adv. Neural Inf. Process. Syst., 2001, 13, pp. 294–300
16 Widrow, B., and Hoff, M.E.: 'Adaptive switching circuits'. WESCON Convention Record, 1960, Vol. IV, pp. 96–104
17 Tang, T.B., Johannessen, E., Wang, L., Astaras, A., Ahmadian, M., Murray, A.F., Cooper, J.M., Beaumont, S.P., Flynn, B.W., and Cumming, D.R.S.: 'Towards a miniature wireless integrated multi-sensor microsystem for industrial and biomedical applications', IEEE Sens. J. Special Issue on Integrated Multisensor Systems and Signal Processing, 2002, 2, (6), pp. 628–635
18 Amari, S.: 'A theory of adaptive pattern classifiers', IEEE Trans. Electron. Comput., 1967, 16, (3), pp. 299–307
19 Barkai, N., Seung, H.S., and Sompolinsky, H.: 'On-line learning of dichotomies', Adv. Neural Inf. Process. Syst., 1994, 7, pp. 303–310
20 Murata, N., Muller, K., Ziehe, A., and Amari, S.: 'Adaptive on-line learning in changing environments', Adv. Neural Inf. Process. Syst., 1996, 9, pp. 599–605
21 Hagan, M.T., Demuth, H.B., and Beale, M.H.: 'Neural network design' (PWS Publishing, Boston, MA, USA, 1996)

34

IEE Proc.-Nanobiotechnol. Vol. 151, No. 1, February 2004

132

Toward a Miniature Wireless Integrated Multisensor Microsystem for Industrial and Biomedical Applications

Tong Boon Tang, Erik A. Johannessen, Lei Wang, Alexander Astaras, Mansour Ahmadian,
Alan F. Murray, *Senior Member, IEEE*, Jonathan M. Cooper, Steve P. Beaumont, Brian W. Flynn, and
David R. S. Cumming, *Member, IEEE*

Abstract—This paper presents our work toward the integration of a multisensor microsystem with wireless communication, using system-on-chip (SoC) methodology. Four different forms of microelectronic sensors have been fabricated on two separate 5×5 mm^2 silicon chips measuring pH, conductivity, dissolved oxygen concentration, and temperature. The sensors are integrated with a sensor fusion chip comprising analog circuitry for sensor operation and signal amplification prior to digital decoding and transmission. The microsystem prototype will be packaged in a miniature capsule, which measures 16 mm $\times$ 55 mm including batteries and dissipates 6.3 mW for a minimal life cycle of 12 h.

Index Terms—Laboratory-on-a-chip (LoC), microsystem, multisensor array, system-on-chip (SoC), wireless communication.

Fig. 1. Integrated multisensor microsystem for the prototype (block diagram).

I. INTRODUCTION

THERE is a considerable interest in the development of ultraminiature and low-power multisensor microsystems for use in applications such as medical diagnosis [1] and environmental monitoring [2]. Industrial applications in water purification, wastewater analysis [3], and radiation control [4], [5] also have technologically attractive prospects. Such systems must exhibit low-cost production, robustness of use, and real-time data processing. While laboratory-on-a-chip (LoC) and system-on-chip (SoC) methodologies are not new, the integration of the two technologies poses many new research challenges. LoC technology offers low-cost, efficient chemical synthesis and the potential for a full-scale biological and clinical analysis on a single, miniaturized device [6]. Meanwhile SoC technology offers an implementation methodology in which a low-power system comprises a set of optimized

Intellectual Property (IP) blocks that can be designed rapidly and economically [7], [8].

The integration of LoC and SoC technologies is aided by improved digital telecommunication coding techniques such as code division multiplexing access (CDMA). In the past, the deployment of conventional distributed monitoring systems was hindered by issues concerning the complex installation of communication systems with respect to interface wiring. However, it is nowadays possible to achieve a usable signal-to-noise ratio (SNR) and thus accurate measurement for a low-power wireless sensor microsystem, clearly within spatial limits [9].

Since the design that forms the essential component of the SoC methodology can easily be integrated with second-generation components, it is anticipated that the novel combination of circuits, sensors, and methodologies embodied in this first implementation will form the basis for future multisensor microsystems.

II. SYSTEM SPECIFICATION

The main features of the microsystem will in the future include an integrated multisensor array, analog signal conditioning circuits, a digital signal processor, a wireless transmitter, and a power source (Fig. 1). For this prototype, both analog signal conditioning circuits and a digital signal processor have been implemented as an application-specific integrated circuit (ASIC) on a silicon chip.

A. Integrated Multisensor

In order to address generic issues known to be of interest to medicine (for example, the detection of abnormalities by

Fig. 2. An optical micrograph of (a) sensor chip 1 comprising the pH, conductivity, and temperature sensor. Sensor chip 2 (b) comprises the oxygen sensor and an optional NiCr resistance thermometer temperature sensor used to measure the temperature of the DO solution.

real-time signal monitoring of the gastro-intestinal (GI) tract [10]), four different types of microelectronic sensors have been employed. The sensor array consists of a dissolved oxygen (DO) sensor [11], [12], a pH-sensitive ion-sel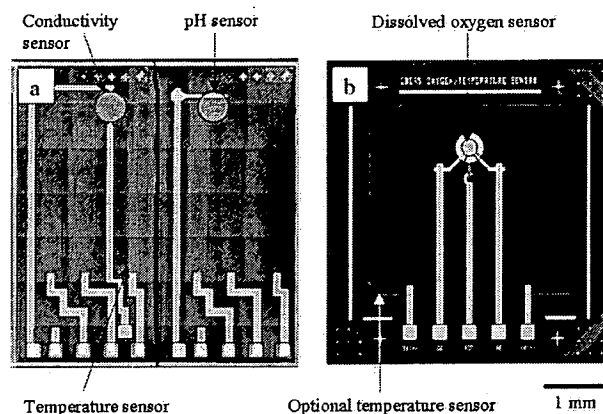ective field effect transistor (ISFET) [13], [14], a standard PN-junction silicon diode temperature sensor [15], and a dual electrode direct contact conductivity sensor [16]. The temperature sensor measures the body core temperature. However, it is also essential for temperature compensation of the other three sensors, since the respective parameters are highly temperature-dependent. The pH sensor measures the acidity of the stomach, although it will also be used to predict the location of the pill within the GI tract, since various portions of the digestive system have an associated range of pH values. The conductivity sensor measures the total level of dissolved solids and thus provides an indirect estimation of the intestinal contents. The oxygen sensor will map the activity of aerobic bacteria within the intestine. The complete sensor array is fabricated on two 5×5 mm^2 silicon chips.

The DO sensor, which includes an optional NiCr resistance thermometer [Fig. 2(b)], was fabricated on a single chip. The sensor consists of a three-electrode electrochemical cell of 500 μm in diameter and a volume of 7.9 nL. The working electrode comprises a microelectrode array of 121 circular gold electrodes, each 5 μm in diameter, to reduce response time and to prevent diffusion-limited transport to the electrode surface. The interelectrode spacing is 20 μm, with a total area of the working electrode of approximately 2.5×10^{-9} m^2. An integrated Ag–AgCl electrode (1.5×10^{-8} m^2) made from 500-nm-thick evaporated silver, of which 50% was oxidized to AgCl in a 1.0 M KCl solution with the aid of chronoamperometry, acts as the reference electrode. The counter electrode is made of gold, covering an area of 1.0×10^{-7} m^2. In future devices, the electrochemical cell will contain a 0.1 M KCl gel (agarose) to promote the constant concentration of chloride ions required to maintain a stable potential of the Ag–AgCl reference electrode and to prevent leakage to the external solution. The electrolyte chamber will be covered by a 20-μm-thick

layer of oxygen-permeable teflon, forming a barrier to the surrounding medium in which the DO is to be measured.

The second chip [Fig. 2(a)] is based on a commercial ISFET and silicon diode temperature sensor (ESIEE, France). The conductivity sensor and a miniaturised Ag–AgCl reference electrode are defined by post lithographic pattern integration. The conductivity sensor comprises two gold electrodes with an area of 5.0×10^{-10} m^2 each spaced 400 μm apart, which operate in a direct contact mode with the surrounding solution in which the conductivity is measured. In contrast, the Ag–AgCl reference electrode (1.5×10^{-8} m^2) is fabricated in a similar manner to the oxygen sensor. Similarly, the ISFET pH sensor and its reference electrode are located in a 7.9-nL electrode chamber designed to contain a 0.1-M KCl gel electrolyte solution to promote a constant concentration of chloride ions in the presence of the Ag–AgCl electrode. In future devices, a final 50-μm-thick layer of cation-selective Nafion® membrane will cover the electrode chamber and forms a physical barrier between the chamber and the solution in which pH is to be measured.

B. Sensor Fusion

The third chip in the microsytem is the ASIC. It includes the analog signal conditioning circuits for each sensor, analog multiplexers, a 10-b analog-to-digital converter (ADC) and a digital-to-analog converter (DAC), an oscillator, as well as digital modules for microsystem scheduling, coding and transmitter control. With minimal compromise in terms of the sensitivity of the sensors, the analog circuits aim to minimize physical dimensions and power consumption. At first glance, some of our design choices would appear at odds with common practice that tends toward far more complex circuits such as the lock-in amplifier or the instrumentation amplifier. However, it is the aim of this project to achieve miniature size and high robustness while keeping power consumption and component count to a minimum—so the common design priorities have been changed.
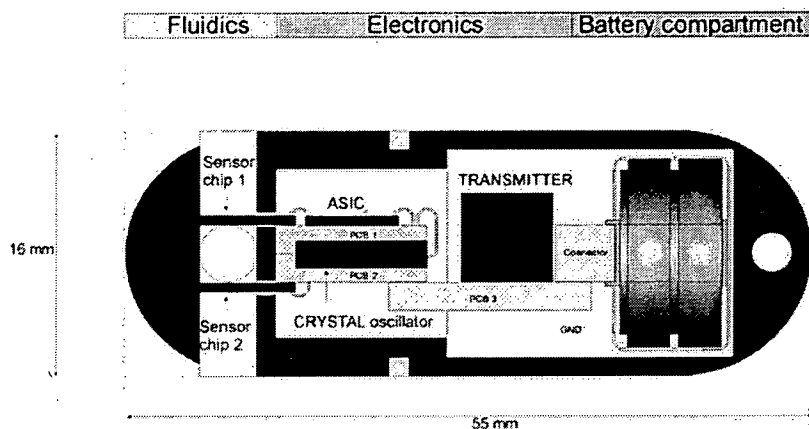
Fig. 3.   A schematic illustrating the design of the future microsystem. The components, which comprise the sensors, ASIC, transmitter, and batteries, are attached to a PCB platform upon integration in a capsule machined from solid PEEK. The capsule has an outer diameter of 16 mm and a length of 55 mm.

In this prototype, the microsystem was designed to work as a real-time multisensor transducer without local signal processing. Although a finite state machine was used for task scheduling in the present configuration, it will be replaced by a small microprocessor for self-reconfiguration in the future. As shown in Fig. 1, this state machine as part of the digital platform is timed from a simple RC relaxation oscillator without temperature compensation. The timing precision of the oscillator is not high, owing to the fabrication tolerance of on-chip resistors and capacitors, but this had been foreseen and therefore we ensured that it would not have any strong adverse effect on our system specifications. Our approach is to mix a relatively low signal bandwidth of 1.51 kHz with the transmitter's reference frequency of 20 MHz so that there is minimal shift (less than 0.01%) in the transmitted signal's frequency. The nominal clock frequency of the RC oscillator is 24.15 kHz with a duty cycle of 49.8%. This low clock frequency provides an adequate sampling rate whilst keeping the dynamic power dissipation acceptably low.

For future implementation, an analog stochastic classifier based on the restricted Boltzmann machine (RBM) [17] will be introduced to achieve intelligent sensor fusion. Although artificial neural networks (ANNs) are well established in many applications, most of the data computation is done through soft computing [18]–[20] and the rest is achieved through either microcontroller units (MCU) or digital signal processing (DSP) that require huge computational power. Unlike the Multi-layer Perceptron (MLP) that requires high accuracy (12 b) in weight memory [21] and a complex, supervised update rule, the RBM design can be implemented easily with a simple weight changing circuit [22]. Moreover, the RBM designs have demonstrated high reliability in heartbeat analysis [22] and hand-writing recognition [23].

C. Power Management

Two smart power management strategies have been implemented. The primary strategy is to disable "idle" function blocks during different task phases. The second strategy uses a simple DSP data compression algorithm to achieve low-power serial bitstream transmission. The algorithm decides when transmission is required by comparing the most recent sample with the previous sampled data. This technique is particularly effective when the measuring environment is at quiescence, a commonly encountered situation in many applications.

The algorithm has been tested using an extended data set, which comprises 67 cases of 24-h dynamic gastric pH data [24]. The raw data have an average kurtosis of 5.3 ± 2.8 and an autocorrelation of 0.99 ± 0.01. The compression ratio, that is defined as the ratio of the raw data stream length to the compressed data stream length, is found to be equal to 2.1 ± 0.1 with a distortion degree of 0.008 ± 0.001 with the given data set. If similar conditions apply to the microsystem discussed in this paper, a total power saving of approximately 50% for the transmitter can be achieved simply by introducing a small digital system overhead ($< 10\%$).

D. Future Wireless Communications

The wireless communication is proposed to be achieved by implementing a simple miniature low-power transmitter that uses a frequency-shift-keying (FSK) modulation method. The transmitter is designed out of ultraminiaturized discrete components, yielding a $5 \times 8 \times 11$ mm^3 package. This transceiver has a crystal oscillator that is bonded onto the substrate of the printed circuit board (PCB) (Fig. 3). This oscillator provides a highly precise carrier frequency that will allow reliable data collection at the base station despite the inaccuracy of the timing of the ASIC. Preliminary experiments have shown that the transmitter has a cover range of 1 m and the received signal has an SNR of 78 dB at 20 MHz. A fully integrated SoC implementation of a transmitter using standard CMOS technology is being developed.
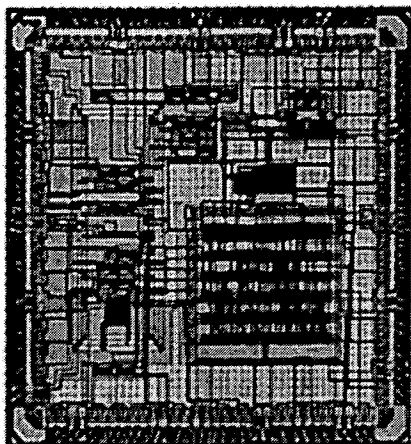
Fig. 5. Sensitivity of the temperature sensor as a function of output voltage over a linear dynamic range from 0 °C to 70 °C ($n = 5$).

Fig. 6. Responsivity of the conductivity sensor versus output voltage over a logarithmic dynamic range from 0.05 to 10 mS·cm^{-1}, performed with prefabricated sensors demonstrating the functionality of the circuit ($n = 5$).

Fig. 4. Optical micrograph of the ASIC comprising both analog and digital systems.

E. Packaging

Interconnection among the sensor chips, the ASIC, and the transmitter were designed for wire bonding to a miniature PCB or by direct connection to adjacent chips. Future implementations will employ more robust packaging techniques such as "flip-chip" bonding. The power source will comprise two standard 26-mAh, 1.5-V silver oxide button cells (Seiko Microsystems, Japan) and the entire proposed prototype microsystem will easily fit into a single 16 mm × 55 mm capsule (Fig. 3) made of polyether-terketone (PEEK).

III. RESULTS AND DISCUSSION

The ASIC was fabricated using a standard 3-V, 2-poly, 3-metal, 0.6-μm CMOS technology process. It was received both as an unpackaged die format for use in the microsystem and in a standard 68-pin package for lab testing purposes (see Fig. 4). The chip is pad-limited with a total area of 4.5 × 4.5 mm^2. There are a total of 16 000 gates, 60% of which are digital while the remaining 40% form the analog system. The ASIC has 64 pads, of which 20 are power pads and 44 are I/O pads. Approximately 50% of the I/O pads are designated for preliminary test purposes of the chip.

Testing of the ASIC and the sensors was carried out on a laboratory test-bench using hard-wired connections to collect data that will be transmitted in a wireless fashion in future iterations of the device. Figs. 5–7 shows in situ test results from the temperature, conductivity, and pH sensor channels that have been completed so far. The correct functionality of the oxygen sensor channel has also been verified. The results are summarized in Table I. The applied voltages are specified from the 0-V reference based on a ±1.5-V power supply. In the worst case scenario whereby all the output data bits are constantly "HIGH," the entire microsystem consumes a total of 2.1 mA of current (refer to Table II), which allows the capsule to continuously operate for more than 12 h using two 26-mAh Ag$_2$O cells as its power supply.
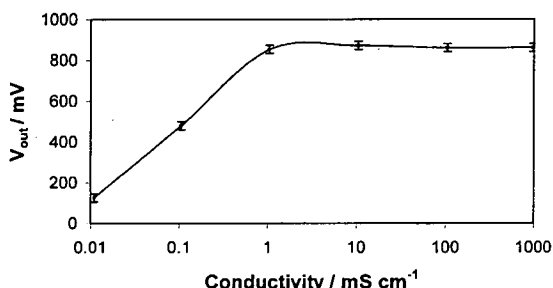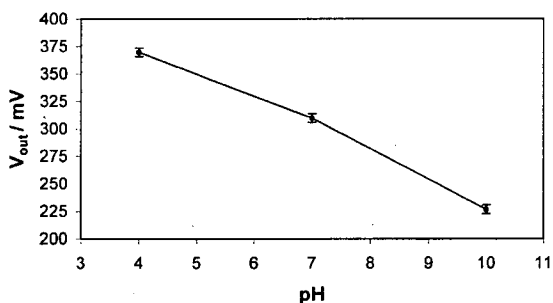
Fig. 7. Sensitivity of the pH sensor as a function of output voltage over a linear dynamic range from pH 4 to 10 ($n = 5$).

Based on the laboratory tests, there is a constant noise level of 3 mV peak-to-peak (equivalent to one least significant bit (LSB) of the ADC) on the ASIC and the sensors. Since the sensors in a harsh environment will be noisier, the second LSB is required to provide an adequate noise margin. Hence, we anticipate that the 10-b ADC has an effective resolution of 8 b.

Temperature Sensor: The silicon diode (PN-junction) temperature sensor operates linearly within the dynamic range from 0 °C to 70 °C (see Fig. 5) with a sensitivity of 31.5 mV °C^{-1} after signal amplification. The forward-biased diode voltage, as a function of temperature, is maintained by a constant bias current of 15 μA.

TABLE I
SUMMARY OF THE SENSORS' PROPERTIES IN CORRESPONDING CHANNELS

Channel	Dynamic range	Cross correlation	Resolution
Temperature	$0 - 70 \,^\circ C$	82.9%	$0.4 \,^\circ C$
Conductivity	$0.05 - 10 \text{ mS} \cdot \text{cm}^{-1}$	99.5%	0.02 mS cm^{-1}
pH	pH 4 – 10	99.3%	0.64 pH unit
Dissolved oxygen	$0 - 8.2 \text{ mg L}^{-1}$	95.0%	0.08 mg L^{-1}

* The resolution is derived based on eight usable bits out of the possible 10 b provided by the ADC.

TABLE II
SUMMARY OF THE WORST CASE SCENARIO OF THE POWER CONSUMPTION
FOR THE ENTIRE MICROSYSTEM WITH A NOMINAL OPERATING
VOLTAGE OF 3 V (± 1.5 V)

Component	Current (mA)	Power (mW)
Analogue (excluding ADC)	0.9	2.7
Digital (including ADC)	0.2	0.6
Transmitter	1.0	3.0
Total	2.1	6.3

* The transmitter is operating at a duty cycle of 15%.

Conductivity Sensor: The conductivity measurement channel of the chip was tested with prefabricated sensors, demonstrating the functionality of the circuit. It exhibited a logarithmic performance from 0.05 to 10 mS·cm^{-1} (Fig. 6), as expected. The voltage amplifier circuit (Fig. 8) on the ASIC forms the electronic interface and measures the impedance, that is inversely proportional to the conductivity, of the sample fluid [device under test (DUT)]. The output voltage saturates at a conductivity above 10 mS·cm^{-1}. However, the formation of an electric double layer (the build up of charged particles on the electrode surface) also affects the sensor's performance. Thus, in order to neutralize the effect of the double layer, the conductivity sensor will be ultimately driven by alternating current. In addition, a dual-electrode, contactless conductivity sensor will be used to reduce the effect of faradaic impedance [25] and to avoid biofouling at the tradeoff of slower time response and poorer detection limits [16].

pH Sensor: The response from the single ISFET pH sensor exhibited linear characteristics over a dynamic range from pH 4 to pH 10 (using standard pH buffers), with an average sensitivity of 23.4 mV pH^{-1} (Fig. 7). However, the stability of the miniaturized reference electrode posed a challenge upon integration with the pH sensor, as any instability of the reference electrode would cause undesirable sensor drift. One possible solution is to combine a sensing ISFET and a reference REFET transistor with two different pH-sensitive layers (e.g., Ta$_2$O$_5$ and oxynitride/Si$_3$N$_4$), respectively [26]. Although the differential sensitivity is small, it can be improved by: 1) reducing the pH sensitivity of the REFET with a polyvinylchloride membrane [27]; 2) increasing the sensitivity of the ISFET by enhancing the proton affinity of the ion-sensitive membrane; or 3) an alternative use of immobilized enzymes on the gate structure of the transducer [28]. However, the differential mode of operation is subject to "crosstalk" between the two transistors if the protons (the by-product of the enzyme reaction) diffuse
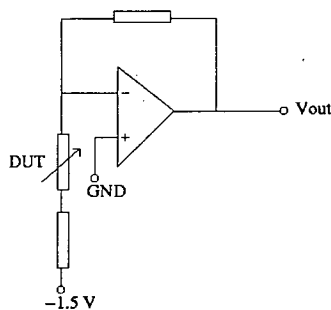
Fig. 8. A voltage amplifier used as an interface circuit for the conductivity sensor. The test fluid is modeled as a resistor (DUT) in the current bias block.
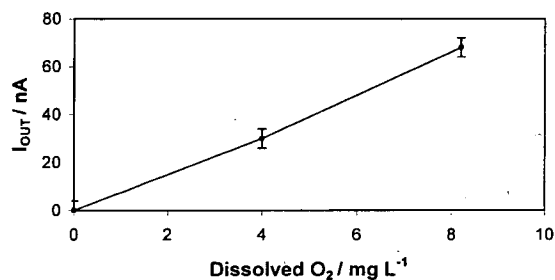
Fig. 9. Sensitivity of the DO sensor versus current over a linear dynamic range from 0 to 8.2 mg L^{-1}, performed with a standard lab potentiostat demonstrating the functionality of the sensor ($n = 5$).

to the REFET, resulting in a false signal. Thus, the single ISFET sensor with an integrated Ag–AgCl reference is still the preferred choice in combination with the simplicity of its electronic interface circuit.

DO Sensor: The DO sensor, which was tested using a standard laboratory potentiostat (Bioanalytical Systems, US), exhibits a linear response with respect to the DO concentration in water over a dynamic range from 0 to 8.2 mg L^{-1} (oxygen saturated water at 25 °C). A sensitivity of 0.12 mg O$_2$ nA^{-1} (Fig. 9) was achieved with an applied working electrode potential of -700 mV with respect to the Ag–AgCl reference. Oxygen saturated water was equilibrated in the lab 24 h prior to the experiment and oxygen was subsequently removed by the chemical titration of the reducing agent Na$_2$SO$_4$. The level of DO in the solution was monitored by a standard Clark O$_2$ electrode (Orion Research Inc, US). The oxygen electrodes were subject to a 2 min O$_2$ plasma clean prior to use, since
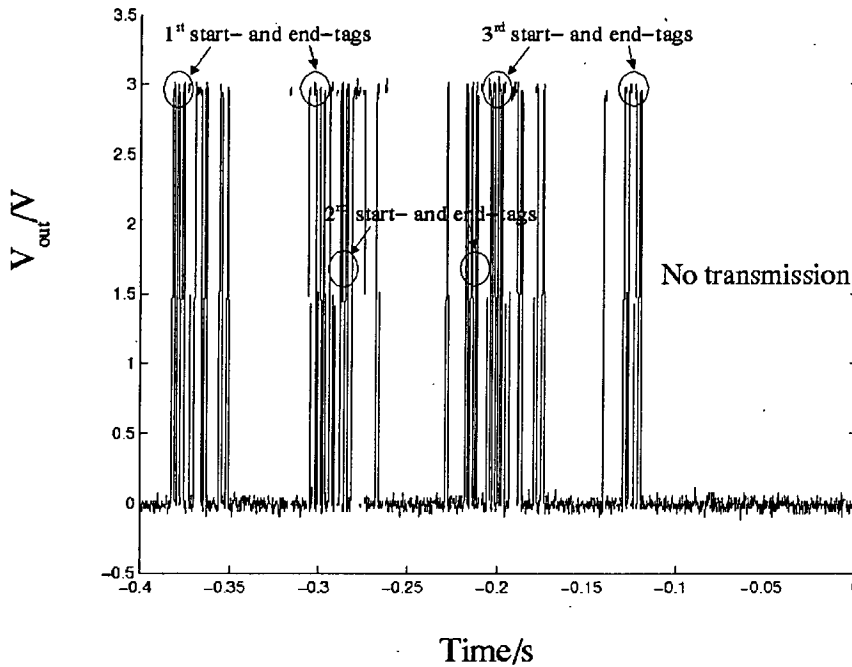
Fig. 10. Output (voltage) of the transmitter controller versus time at an operating frequency of 24.15 kHz. The output signal is used to modulate the transmitter carrier frequency, based on the frequency-shift-keying modulation method.

any residue on the electrode surface after fabrication degrade sensor performance seriously in the electrochemical measurement system.

Transmission Control: Each packet of transmission data will consist of a start-tag (8-b 10 101 010), four sets of sensor data (4 × 10-b), 8-b data compression overhead and an end-tag (8-b 01 010 101) in sequence. Each packet will be transmitted three times to improve the reliability of the transmission. As highlighted in Fig. 10, the start- and end-tags have been successfully identified from the bitstream recorded at the transmission control node. This bitstream will be fed into the transmitter to modulate the transmitter carrier frequency. Besides, it is observed that there is no more transmission after the first three packets since the test environment remains at quiescent. This has hence demonstrated the functionality of the compression algorithm.

IV. CONCLUSION

An ASIC for a multisensor microsystem has been designed and implemented with a microelectronic sensor array and a direct communication link. All sensor and circuit components operate as intended and are designed to be amenable to SoC integration. An SoC design methodology has also been developed. Future implementations will lead to further miniaturization, wireless data transmission, and a more sophisticated system specification including greater sensor diversity, sensor fusion, and dynamic reconfigurability.

REFERENCES

[1] T. Akin, K. Najafi, and P. Bradley, "A wireless implantable multichannel digital neural recording system for a micromachined sieve electrode," *IEEE J. Solid-State Circuits*, vol. 33, pp. 109–118, Jan. 1998.

[2] C. Grimes, M. Jain, R. S. Singh, Q. Cai, A. Mason, K. Takahata, and Y. Gianchandani, "Magnetoelastic microsensors for environmental monitoring," in *Proc. 14th IEEE Int. Conf. Micro Electro Mechanical Systems*, 2001, pp. 278–281.

[3] W. Bourgeois, J. E. Burgess, and R. M. Stuetz, "On-line monitoring of wastewater quality: A review," *J. Chem. Technol. Biotechnol.*, vol. 76, no. 4, pp. 337–348, 2001.

[4] M. Muller, W. Budde, R. Gottfried, A. Hubel, R. Jahne, and H. Kuck, "A thermoelectric infrared radiation sensor with monolithically integrated amplifier stage and temperature sensor," *Sens. Actuators A, Phys.*, vol. 54, no. 1–3, pp. 601–605, 1996.

[5] E. G. Moreno, B. Iniguez, M. Roca, J. Segura, and S. Sureda, "CMOS radiation sensor with binary output," *IEEE Trans. Nucl. Sci.*, vol. 42, pp. 174–178, 1995.

[6] P. Wilding and L. J. Kricka, "Micro-microchips: Just how small can we go?," *Trends Biotechnol.*, vol. 17, pp. 465–468, 1999.

[7] B. Clement, R. Hersemeule, E. Lantreibecq, B. Ramanadin, P. Coulomb, and F. Pogodalla, "Fast prototyping: A system design flow applied to a complex system-on-chip multiprocessor design," in *Proc. Design Automation Conf.*, 1999, pp. 420–424.

[8] Y. H. Lee, K. W. Kwon, J. T. Kim, and C. D. Lee, "IP development and management of IP DB enabling efficient system-on-chip design," in *Proc. 1st IEEE Asia Pacific Conf. ASIC's*, 1999, pp. 229–232.

[9] A. Mason, N. Yazdi, A. V. Chavan, K. Najafi, and K. D. Wise, "A generic multielement microsystem for portable wireless applications," *Proc. IEEE*, vol. 86, pp. 1733–1746, Aug. 1998.

[10] L. Wang, T. B. Tang, E. A. Johannessen, A. Astaras, M. Ahmadian, A. F. Murray, J. M. Cooper, S. P. Beaumont, and D. R. S. Cumming, "Integrated micro-instrumentation for dynamic gastro-intestinal tract monitoring," in *Proc. 19th IEEE IMTC*, 2002, pp. 1717–1720.

[11] L. C. Clark Jr. and C. Lyons, "Electrode systems for continuous monitoring in cardiovascular surgery," *Annu. NY Acad. Sci.*, vol. 102, pp. 29–45, 1962.

[12] G. Jobst, G. Urban, A. Jachimowicz, F. Kohl, and O. Tilado, "Thin film clark-type oxygen sensor based on novel polymer membrane systems for in vivo and biosensor applications," *Biosens. Bioelectron.*, vol. 8, pp. 123–128, 1993.

[13] N. Jaffrezic-Renault and C. Martelet, "Semiconductor-based micro biosensors," *Synthetic Metals*, vol. 90, pp. 205–210, 1997.

[14] P. Bergveld, "Development, operation and application of the ion-sensitive field effect transistor as a tool for electrophysiology," *IEEE Trans. Biomed. Eng.*, vol. BME-19, pp. 342–351, 1972.

[15] M. H. Rashid, *Microelectronics Circuit Analysis and Design*. Boston, MA: PWS, 1999.

[16] G. Weber, M. Johnck, D. Siepe, A. Neyer, and R. Hergenroder, "Capillary electrophoresis with direct and contactless conductivity detection on a polymer microchip," in *Micro Total Analysis Systems 2000*. Amsterdam, The Netherlands: Kluwer, 2000.

[17] Y. Freud and D. Haussler, "Unsupervised learning of distributions on binary vectors using two layer network," *Adv. Neural Inform. Process. Syst.*, vol. 4, pp. 912–919, 1991.

[18] A. Charef, A. Ghauch, P. Baussand, and M. Martin-Bouyer, "Water quality monitoring using a smart sensing system," *Measurement*, vol. 28, pp. 219–224, 2000.

[19] P. E. Keller, R. T. Kouzes, and L. J. Kangas, "Three neural network based sensor systems for environmental monitoring," in *Proc. IEEE Int. Conf. Electro*, 1994, pp. 377–382.

[20] B. R. Bracio, W. Horn, and D. P. F. Moller, "Sensor fusion in biomedical systems," in *Proc. IEEE EMBS Conf.*, vol. 3, 1997, pp. 1387–1390.

[21] P. Moerland and E. Fiesler, "Neural network adaptations to hardware implementations," in *Handbook of Neural Computation*, 1997, ch. E1.2.

[22] P. Fleury, R. J. Woodburn, and A. F. Murray, "Matching analogue hardware with applications using the products of experts algorithm," in *Proc. IEEE Eur. Symp. Artificial Neural Networks*, 2001, pp. 63–67.

[23] G. Mayraz and G. E. Hinton, "Recognizing hand-written digitals using hierarchical products of expects," *Adv. Neural Inform. Process. Syst.*, vol. 13, pp. 953–959, 2000.

[24] L. Wang, C. Zheng, and S. Guo, "Development of new ambulatory monitor for multi-channel gastroesophageal pH combining with cardiogram," *Chin. J. Med. Instrum.*, vol. 22, no. 5, pp. 259–263, 1998.

[25] F. Laugere, G. W. Lubking, A. Berthold, J. Bastemeijer, and M. J. Vellekoop, "Downscaling aspects of a conductivity detector for application in on-chip capillary electrophoresis," *Sens. Actuators A, Phys.*, vol. 92, pp. 109–114, 2001.

[26] D. Wilhelm, H. Voigt, W. Treichel, R. Ferretti, and S. Prasad, "pH sensor based on differential measurements on one pH-FET chip," *Sens. Actuators B, Chem.*, vol. 4, pp. 145–149, 1991.

[27] A. Errachid, J. Bausells, and N. Jaffrezic-Renault, "A simple REFET for pH detection in differential mode," *Sens. Actuators B, Chem.*, vol. 60, pp. 43–48, 1999.

[28] S. V. Dzyadevycha, T. M. Anhb, A. P. Soldatkina, N. D. Chien, N. Jaffrezic-Renault, and J. M. Chovelon, "Development of enzyme biosensor based on pH-sensitive field-effect transistors for detection of phenolic compounds," *Biochemistry*, vol. 55, no. 1–2, pp. 79–81, 2002.

Erik A. Johannessen received the B.Sc. degree in natural sciences from the University of Tromsø, Norway, in 1997 and the Ph.D. degree from the University of Liverpool, Liverpool, U.K., in 2002.

The Ph.D. project, which was a collaboration between the Universities of Liverpool and Glasgow, was funded by the Wellcome Trust focusing on the development of ultrasmall nanocalorimetric sensors for integration in high-density assay screening of cells in medical research. He is currently a Research Assistant with the Department of Electronic and Electrical Engineering, University of Glasgow, Glasgow, U.K.. His current field of research interest is in bio and nanoelectronics, silicon microfabrication, and the development of mobile analytical microsystems.

Lei Wang received the B.S. degree in information and control engineering and the Ph.D. degree in biomedical engineering from Xi'an Jiaotong university, Xi'an, China, in 1995 and 2000, respectively.

After graduation, he was an academic visitor at the Department of Mechanical Engineering, University of Dundee, Dundee, U.K. In 2001, he joined the Department of Electronics and Electrical engineering, University of Glasgow, Glasgow, U.K., as a Post-Doctoral Researcher. His research interests focus on physiological measurement, digital signal processing, and integrated circuit design.

Alexander Astaras was born in Thessaloniki, Greece, in 1974. He received the B.A. degree in physics from Oberlin College, Oberlin, OH, in 1995. He is currently working toward the Ph.D. degree in electronics engineering from the University of Edinburgh, Edinburgh, U.K., on analog VLSI circuit design for unsupervised probabilistic neural networks.

He also works as a Research Assistant for the Institute for System Level Integration in Livingston, U.K., on system-on-chip integration for wireless microinstrumentation. His research interests revolve around biomedical diagnostics, wireless microinstrumentation, hardware implementations of artificial neural networks, and mixed signal system-on-chip VLSI design.

Tong Boon Tang received the B.Eng. (Hons) degree in electronics and electrical engineering (communications) from The University of Edinburgh, Edinburgh, U.K., in 1999. He is currently working toward the Ph.D. degree in intelligent sensor fusion at the same university.

He then worked as an ASIC Engineer with Singapore Design Centre, Lucent Technologies. In 2001, he joined the Integrated System Group at The University of Edinburgh as a Research Associate. His research interests lie in unsupervised stochastic neural algorithms, analog VLSI design, and smart sensors, especially in data fusion.

Mansour Ahmadian received the B.S. degree in electronic engineering in 1989 from Khajeh-Nasir-Toosi University, Iran, the M.Sc. degree in biomedical engineering in 1992 from Sharif University, Iran, and the Ph.D. degree in biomedical engineering in 2002 from The University of Edinburgh, Edinburgh, U. K.

He is currently doing his post-doctorate research at the University of Edinburgh, Edinburgh, U.K., in the School of Engineering and Electronics. His research interests are medical signal and image processing, parallel processing, communication systems and microwave.
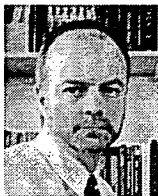
Alan F. Murray (M'91–SM'93) was born in 1953 in Edinburgh, U.K., He received the B.Sc. degree (Hons.) in physics and the Ph.D. degree in solid-state physics from the University of Edinburgh in 1975 and 1978, respectively.

He worked for three years as a Research Physicist (two in Canada) and for three years as an Integrated Circuit Design Engineer. In 1984, he was appointed a Lecturer in Electrical Engineering at Edinburgh University and became a Reader in 1991 and a Professor of Neural Electronics in 1994. He is interested in all aspects of neural computation, and hardware issues and applications have been his primary research interest since 1985. In 1986, he developed the "pulse stream" method for neural integration. His interests have since widened to include all aspects of neural computation, particularly hardware-compatible learning schemes, probabilistic neural computation and neural forms that utilize the temporal- and noisy characteristics of analogue VLSI—as well as applications of hardware neural networks. He is also developing a new interest in the interface between silicon and neurobiology, along with colleagues in Biomedical Sciences and in Glasgow University. He has over 190 publications, including an undergraduate textbook and research texts on neural VLSI, applications of neural networks and noise in neural training (with Peter Edwards).

Dr. Murray is a fellow of the Institution of Electrical Engineers, a member of INNS, and a fellow of the Royal Society of Edinburgh.

Steve P. Beaumont received the M.A. and Ph.D. degrees from University of Cambridge, U.K. in 1975 and 1979, respectively.

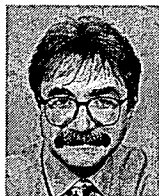He is Director of the Institute for System Level Integration (ISLI), Livingston, U. K., and Professor of Nanoelectronics at the University of Glasgow, Glasgow, U.K. His research interests have covered nanofabrication technology, fabrication of deep submicrometer and nanometer-scale devices and millimeter-wave integrated circuit design and fabrication. He has published over 200 technical papers in these areas. Having founded ISLI in 1998, his interests now lie in the design and applications of systems-on-chip, including integration with advanced peripheral components such as sensors and laboratory-on-a-chip devices.

Brian W. Flynn received the B.Sc. and Ph.D. degrees in electrical engineering from the University of Edinburgh, Edinburgh, U.K.

He is presently a Senior Lecturer in the Department of Electrical Engineering,University of Edinburgh. After a period working on a Research Fellowship at the University of Edinburgh, he moved to Marconi Communications Systems, Ltd., as a Development Engineer working on microwave radio link systems. In 1980, he was appointed to a lectureship in Electrical Engineering at the University of Edinburgh and has taught courses in electromagnetics and RF/microwave design. He is the coauthor of a textbook on the design of switched-mode power supplies and has worked on the design of switched-mode power supplies at high switching speeds.

Jonathan M. Cooper received the B.Sc. degree in biological sciences from the University of Southampton, Southampton, U.K., in 1983 and the Ph.D. degree in sensor technology from the University of Cranfield, Cranfield, U. K., in 1989.

He is a Professor of Bioelectronics in the Department of Electronic and Electrical Engineering, University of Glasgow, Glasgow, U.K. His current field of research is in medical diagnostics, thermal, electrochemical, and biochemical detection methods, lithographic patterning, and silicon microfabrication design for the development of micro total analysis systems.

David R. S. Cumming (M'97) received the B.Eng. degree from the University of Glasgow, Glasgow, U.K., in 1989 and the Ph.D. degree from the University of Cambridge, Cambridge, U. K., in 1993.

He has worked variously on mesoscopic device physics, RF characterization of novel devices, fabrication of diffractive optics for optical and submillimeter wave applications, and microelectronic design. He is presently a Senior Lecturer and EPSRC Advanced Research Fellow in Electronics and Electrical Engineering at the University of Glasgow, where he leads the Microsystem Technology Group.

References

[1] P. E. Allen and D. R. Holberg, *CMOS Analog Circuit Design*, ch. 5, pp. 219–226. New York, USA: Oxford University Press, 1987. Current Sinks and Sources.

[2] S. Middelhoek and A. C. Hoogerwerf, "Smart Sensors: When and Where?," *Sensors and Actuators*, vol. 8, pp. 39–48, 1985.

[3] W. H. Ko and C. D. Fung, "VLSI and intelligent transducers," *Sensors and Actuators*, vol. 2, pp. 239–250, 1982.

[4] K. D. Wise, "Integrated microsystems: Merging MEMS, micropower electronics, and wireless commnunications," in *Proceedings of IEEE ASIC/SoC Conference*, pp. xxiii–xxix, 1999.

[5] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, pp. 51–58, May 2000.

[6] R. C. Luo, C. C. Yih, and K. L. Su, "Multisensor fusion and integration: Approachs, applications, and future research directions," *IEEE Sensors Journal*, vol. 2, pp. 107–119, April 2002.

[7] B. A. Warneke, M. D. Scott, B. S. Leibowitz, L. Zhou, C. L. Bellew, J. A. Chediak, J. M. Kahn, B. E. Boser, and K. S. J. Pister, "An autonomous $16mm^3$ solar-powered node for distributed wireless sensor networks," in *Proceedings of IEEE Sensors*, (Orlando, FL, USA), pp. 1510–1515, June 2002.

[8] J. M. Rabaey, M. J. Ammer, J. L. da Silva Jr., D. Patel, and S. Roundy, "PicoRadio supports ad hoc ultra-low power wireless networking," *Computer*, vol. 33, pp. 42–48, July 2000.

[9] D. Figeys and D. Pinto, "Lab-on-a-chip: A revolution in biological and medical sciences," *Analytical Chemistry*, vol. 72, pp. 330A–335A, May 2000.

[10] G. Martin and H. Chang, "System-on-chip design," in *Proceedings of International Conference on ASIC*, (Shanghai, China), pp. 12–17, October 2001.

[11] G. I. Ltd, "Expanding the scope of GI," April 2006. URL: http://www.givenimaging.com/Cultures/en-US/Given/English.

[12] K. Hirai, M. Hirose, Y. Haikawa, and T. Takenaka, "The development of Honda humanoid robot," in *IEEE International Conference Proceedings on Robotics and Automation*, vol. 2, (Leuven, Belgium), pp. 1321–1326, May 1998.

[13] W. Burgard, A. B. Cremers, D. Fox, D. Hahnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "Experiences with an interactive museum tour-guide robot," *Artificial Intelligence*, vol. 114, pp. 3–55, October 1999.

[14] C. E. Lin and K. L. Chen, "Automated air defense system using knowledge-based system," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 27, pp. 118–124, January 1991.

[15] D. L. Hall, R. J. Linn, and J. Llinas, "A survey of data fusion systems," in *SPIE Conference Proceedings on Data Structure and Target Classification*, vol. 1470, (Orlando, FL, USA), pp. 13–36, April 1991.

[16] R. R. Murphy, "Biological and cognitive foundations of intelligent sensor fusion," *IEEE Transactions on Systems, Man and Cybernetics - Part A: Sytems and Humans*, vol. 26, pp. 42–51, January 1996.

[17] H. E. Stephanou and A. P. Sage, "Perspectives on imperfect information processing," *IEEE Transactions on System, Man, Cybernetics*, vol. 17, pp. 780–798, September 1987.

[18] X. Xu and J. W. Hines, "Sensor calibration and fault detection using neural network techniques," in *Proceedings of Maintenance and Reliability Conference*, (Knoxville, TN, USA), May 1997.

[19] S. Shafer, A. Stentz, and C. Thorpe, "An architecture for sensor fusion in a mobile robot," in *IEEE International Conference Proceedings on Robotics and Automation*, vol. 3, (San Francisco, CA, USA), pp. 2002–2011, April 1986.

[20] Q. Bai and K. D. Wise, "Single-unit neural recording with active microelectrode arrays," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 911–920, August 2001.

[21] H. Yu and K. Najafi, "Circuitry for a wireless microsystem for neural recording microprobes," in *IEEE International Conference Proceedings of Engineering in Medicine and Biology Society*, vol. 1, (Istanbul, Turkey), pp. 761–764, October 2001.

[22] A. F. P. van Putten, M. J. A. M. van Putten, M. H. P. M. van Putten, and P. F. A. M. van Putten, "Multisensor microsystem for pulmonary function diagnostics," *IEEE Sensors Journal*, vol. 2, pp. 636–643, December 2002.

[23] M. Seyama, I. Sugimoto, and T. Miyagi, "Application of an array sensor based on plasma-deposited organic film coated quartz crystal resonators to monitoring indoor volatile compounds," *IEEE Sensors Journal*, vol. 2, pp. 422–427, October 2002.

[24] C. Pohl and J. L. van Genderen, "Multisensor image fusion in remote sensing: concepts, methods and applications," *International Journal of Remote Sensing*, vol. 19, pp. 823–854, March 1998.

[25] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, pp. 1350–1359, May 1999.

[26] R. C. Luo and M. G. Kay, "A tutorial on multisensor integration and fusion," in *Proceedings of IEEE Industrial Electronics Society*, pp. 707–722, November 1990.

[27] P. S. Maybeck, *Stochastic Models, Estimation, and Control*, vol. 1, ch. 1: Introduction. Academic Press Inc., 1979.

[28] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, ch. 7.5: The Kalman Filter, pp. 438–443. Addison Wesley Longman, second ed., 1994.

[29] K. Tanaka, M. Sano, S. Ohara, and M. Okudaira, "A parametric template method and its application to robust matching," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, (Hilton Head Island, South Carolina, USA), pp. 620–627, June 2000.

[30] R. Beale and T. Jackson, *Neural Computing: An Introduction*, ch. 2: Pattern Recognition, pp. 15–37. Institute of Physics, 2001.

[31] P. Smets, "The combination of evidence in the transferable belief model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 447–458, May 1990.

[32] L. A. Zadeh, "Fuzzy logic," *IEEE Computer*, vol. 21, pp. 83–93, April 1988.

[33] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 1998.

[34] P. E. Keller, R. T. Kouzes, and L. J. Kangas, "Three neural network based sensor systems for environmental monitoring," in *Proceedings of the IEEE Electro*, (Boston, MA, USA), pp. 378–382, May 1994.

[35] C. D. Natale, F. A. M. Davide, and A. D'Amico, "A self-organizing system for pattern classification: time varying statistics and sensor drift effects," *Sensors and Actuators B*, vol. 26-27, pp. 237–241, 1995.

[36] P. Wide, F. Winquist, P. Bergsten, and E. M. Petriu, "The human-based multisensor fusion method for artificial nose and tongue sensor data," *IEEE Transactions on Instrumentation and Measurement*, vol. 47, pp. 1072–1077, October 1998.

[37] S. Marco, A. Ortega, A. Pardo, and J. Samitier, "Gas identification with tin oxide sensor array and self-organizing maps: Adaptive correction of sensor drifts," *IEEE Transactions on Instrumentation and Measurement*, vol. 47, pp. 316–321, February 1998.

[38] J. W. Gardner, E. L. Hines, F. Molinier, P. N. Bartlett, and T. T. Mottram, "Prediction of health of dairy cattle from breath samples using neural network with parametric model of dynamic response of array of semiconducting gas sensors," *IEE Proceedings on Sci. Meas. Technology*, vol. 146, pp. 102–106, March 1999.

[39] T. Roppel, D. Wilson, K. Dunman, V. Becanovic, and M. L. Padgett, "Design of a low-power, portable sensor system using embedded neural networks and hardware preprocessing," in *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 142–145, 1999.

[40] B. G. Kermani, S. S. Schiffman, and H. T. Nagle, "Using neural networks and genetic algorithms to enhance performance in an electronic nose," *IEEE Transactions on Biomedical Engineering*, vol. 46, pp. 429–439, April 1999.

[41] F. Sarry and M. Lumbreras, "Gas discrimination in an air-conditioned system," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, pp. 809–812, August 2000.

[42] T. Sundic, S. Marco, J. Samitier, and P. Wide, "Electronic tongue and electronic nose data fusion in classification with neural networks and fuzzy logic based models," in *Proceedings of the IEEE IMTC*, vol. 3, pp. 1474–1479, May 2000.

[43] H. W. Shin, E. Llober, J. W. Gardner, E. L. Hines, and C. S. Dow, "Classification of the strain and growth phase of Cyanobacteria in potable water using an electronic nose system," *IEE Proceedings on Sci. Meas. Technology*, vol. 147, pp. 158–164, July 2000.

[44] M. Lindquist and P. Wide, "Virtual water quality tests with an electronic tongue," in *Proceedings of the IEEE IMTC*, vol. 2, pp. 1320–1324, May 2001.

[45] I. Sayago, M. d. C. Horrillo, S. Baluk, M. Aleixandre, M. J. Fernandez, L. Ares, M. Garcia, J. P. Santos, and J. Gutierrez, "Detection of toxic gases by a tin oxide multisensor," *IEEE Sensors Journal*, vol. 2, pp. 387–393, October 2002.

[46] S. Bermejo, G. Bedoya, V. Parisi, and J. Cabestany, "An on-line water monitoring system using a smart ISFET array," in *Proceedings of the IEEE Conference on Industrial Electronics Society*, pp. 2797–2802, 2002.

[47] G. Bedoya, C. Jutten, S. Bermejo, and J. Cabestany, "Improving semiconductor-based chemical sensor arrays using advanced algorithms for blind source separation," in *Proceedings of the IEEE Sensors for Industry Conference*, (New Orleans, Louisiana, USA), pp. 149–154, January 2004.

[48] M. L. Rodriguez-Mendez, A. A. Arrieta, V. Parra, A. Bernal, A. Vegas, S. Villanueva, R. Gutierrez-Osuna, and J. A. de Saja, "Fusion of three sensory modalities for the multimodal characterization of red wines," *IEEE Sensors Journal*, vol. 4, pp. 348–354, June 2004.

[49] J. R. Stetter and W. R. Penrose, "The electrochemical nose." URL:http://electrochem.cwru.edu/ed/encycl/art-n01-nose.htm, August 2001.

[50] Wikipedia.org, "Principal components analysis," September 2005. URL:http://en.wikipedia.org/wiki/Principal_components_analysis.

[51] K. J. Lang, A. H. Waibel, and G. E. Hinton, "A time-delay neural network architecture for isolated word recognition," *Neural Networks*, vol. 3, no. 1, pp. 23–43, 1990.

[52] S. Sarkaria, "Catastrophic interference." http://www.ee.ubc.ca/ elec592/PDFfiles /Catastrophic_Learning.pdf, 2004.

[53] D. W. Clarke, "Sensor, actuator and plant validation," *IEE Colloquium on Intelligent and Self-Validating Sensors*, pp. 1–8, June 1999.

[54] A. Steinhage and C. Winkel, "A robust self-calibrating data fusion architecture," in *Proceedings of IEEE National Geoscience and Remote Sensing Symposium*, (Honolulu, HI, USA), pp. 963–965, July 2000.

[55] C. S. Tsai, C. C. Tong, and L. E. Oh, "Sensor data correction with neural network incorporating fuzzy logic," in *Proceedings of IEEE International Fuzzy Systems Conference*, (Seoul, Korea), pp. 66–71, August 1999.

[56] N. L. Bris and D. Birot, "Automated pH-ISFET measurements under hydrostatic pressure for marine monitoring application," *Analytica Chimica Acta*, vol. 356, pp. 205–215, 1997.

[57] R. L. Smith and D. C. Scott, "An integrated sensor for electrochemical measurements," *IEEE Transactions on Biomedical Engineering*, vol. 33, pp. 83–90, February 1986.

[58] S. Jamasb, S. D. Collins, and R. L. Smith, "Correction of instability in Ion-selective Field Effect Transistors for accurate continuous monitoring of pH," in *Proceedings of IEEE International Conference of EMBS*, (Chicago, IL, USA), pp. 2337–2340, October/November 1997.

[59] M. Holmberg, F. A. M. Davide, C. D. Natale, A. D'Amico, F. Winquist, and I. Lundstrom, "Drift counteraction in odour recognition applications: lifelong calibration method," *Sensors and Actuators B*, vol. 42, pp. 185–194, 1997.

[60] G. G. Yen and W. Feng, "Winner take all experts network for sensor validation," in *Proceedings of the IEEE International Conference on Control Applications*, (Anchorage, Alaska, USA), pp. 92–97, September 2000.

[61] J. Hendrikse, W. Olthuis, and P. Bergveld, "A method of reducing oxygen induced drift in iridium oxide pH sensors," *Sensors and Actuators B*, vol. 53, pp. 97–103, 1998.

[62] A. Errachid, J. Bausells, and N. Jaffrezic-Renault, "A simple REFET for pH detection in differential mode," *Sensors and Actuators B*, vol. 60, pp. 43–48, 1999.

[63] M. Grattarola, G. Massobrio, and S. Martinoia, "Modelling H^+-Sensitive FET's with SPICE," *IEEE Transactions on Electron Devices*, vol. 39, pp. 813–819, April 1992.

[64] S. Jamasb, S. D. Collins, and R. L. Smith, "A physical model for threshold voltage instability in Si_3N_4-Gate H^+-Sensitive FET's (pH-ISFET's)," *IEEE Transactions on Electron Devices*, vol. 45, pp. 1239–1245, June 1998.

[65] S. Jamasb, "An analytical technique for counteracting drift in ion-selective field effect transistor (ISFETs)," *IEEE Sensors Journal*, 2004.

[66] A. Sachenko, V. Kochan, V. Turchenko, K. Tsahouridis, and T. Laopoulos, "Error compensation in an intelligent sensing instrumentation system," in *Proceedings of IEEE Instrumnetation and Measurement Technology Conference,* (Budapest, Hungary), pp. 869–874, May 2001.

[67] A. A. Platonov, J. Szabatin, and K. Jedrzejewski, "Optimal synthesis of smart measurement systems with adaptive correction of drifts and setting errors of the sensor's working point," *IEEE Transactions on Intrumentation and Measurement,* vol. 47, pp. 659–665, June 1998.

[68] T. Artursson, T. Eklov, I. Lundstrom, P. Martensson, M. Sjostrom, and M. Holmberg, "Drift correction for gas sensors using multivariate methods," *Journal of Chemometrics,* vol. 14, pp. 711–723, 2000.

[69] S. Holmin, C. Krantz-Rulcker, I. Lundstrom, and F. Winquist, "Drift correction of electronic tongue responses," *Institute of Physics Measurement Science Technology,* vol. 12, pp. 1348–1354, 2001.

[70] B. Lazzerini and F. Marcelloni, "Counteracting drift of olfactory sensors by appropriately selecting features," *IEE Electronics Letters,* vol. 36, pp. 509–510, March 2000.

[71] D. Chung and F. L. Merat, "Neural network based sensor array signal processing," in *Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems,* (Washington, DC, USA), pp. 757–764, December 1996.

[72] S. Park and C. S. G. Lee, "Fusion-based sensor fault detection," in *Proceedings of IEEE International Symposium on Intelligent Control,* (Chicago, IL, USA), pp. 156–161, August 1993.

[73] T. H. Guo and J. Nurre, "Sensor failure detection and recovery by neural networks," in *Proceedings of IJCNN,* vol. 1, (Seattle, WA, USA), pp. 221–226, July 1991.

[74] J. C. Seiter and M. D. DeGrandpre, "Redundant chemical sensors for calibration-impossible applications," *Talanta,* pp. 99–106, 2001.

[75] P. Ienne, T. Cornu, and G. Kuhn, "Special-purpose digital hardware for neural networks: An architectural survey," *Journal of VLSI Signal Processing Systems,* vol. 13, pp. 5–25, 1996.

[76] H. C. Card, D. K. McNeill, and C. R. Schneider, "Analog VLSI circuits for competitive learning networks," *Analog Integrated Circuits and Signal Processing,* vol. 15, pp. 291–314, 1998.

[77] R. Woodburn and A. F. Murray, "Implementing artificial neural networks in analogue VLSI," in *Proceedings of the International Conference on Neural Information Processing,* (Dunedin, New Zealand), pp. 658–661, 1997.

[78] R. Coggins, M. Jabri, B. Flower, and S. Pickard, "A hybrid analog and digital VLSI neural network for intracardiacmorphology classification," *IEEE Journal of Solid-States Circuits*, vol. 30, pp. 542–550, May 1995.

[79] P. H. W. Leong and M. A. Jabri, "A low power trainable analogue neural network classifier chip," in *Proceedings of the IEEE Custom Integrated Circuits Conference*, (San Diego, CA, USA), pp. 451–454, May 1993.

[80] K. Asanovic and N. Morgan, "Experimental determination of precision requirements for back-propagation training of artificial neural networks," in *Proceedings of International Conference on Microelectronics for Neural Network*, (Munich, Germany), pp. 9–15, October 1991.

[81] M. Jabri and B. Flower, "Weight perturbation: An optimal architecture and learning technique for analog VLSI feedforward and recurrent multilayer networks," *IEEE Transactions on Neural Networks*, vol. 3, pp. 154–157, January 1992.

[82] J. L. Holt and J. N. Hwang, "Finite precision error analysis of neural network hardware implementations," *IEEE Transactions on Computers*, vol. 42, no. 3, pp. 281–290, 1993.

[83] D. J. Mayes, A. Hamilton, A. F. Murray, and H. M. Reekie, "A pulsed VLSI radial basis function chip," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, vol. 3, (Atlanta, GA, USA), pp. 297–300, May 1996.

[84] P. Moerland and E. Fiesler, *Handbook of Neural Computation*, ch. Chapter E1.2: Neural Network Adaptations to Hardware Implementations. New York, USA: Institute of Physics Publishing and Oxford University Publishing, 1996.

[85] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, *Learning Internal Representations by Error Propagation*, vol. 1 of *Computational models of cognition and perception*, ch. 8, pp. 319–362. Cambridge, MA, USA: MIT Press, 1986.

[86] K. Nishizawa and Y. Hirai, "Hardware implementation of PCA neural network," in *Proceedings of ICONIP*, (Kitakyushu, Japan), pp. 85–88, October 1998.

[87] D. Macq, M. Verleysen, P. Jespers, and J.-D. Legat, "Analog implementation of a kohonen map with on-chip learning," *IEEE Transactions on Neural Networks*, vol. 4, pp. 456–461, May 1993.

[88] D. Hsu, M. Figueroa, and C. Diorio, "Competitive learning with floating-gate circuits," *IEEE Transactions on Neural Networks*, vol. 13, pp. 732–744, May 2002.

[89] T. Higuchi, T. Furuya, K. Handa, N. Takahashi, H. Nishiyama, and A. Kokubu, "IXM2: A parallel associative processor," in *Proceedings of the international symposium on Computer architecture*, (Toronto, Ontario, Canada), pp. 22–31, 1991.

[90] L. O. Chua and T. Roska, "The CNN paradigm," *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 40, pp. 147–156, March 1993.

[91] J. Alspector and B. G. amd R. B. Allen, "Performance of a stochastic learning microchip," in *Advances in Neural Information Processing Systems*, vol. 1, pp. 748–760, 1989.

[92] D. Heckerman, *Learning in graphical models*, ch. A tutorial on learning with Bayesian networks, pp. 301–354. Cambridge, MA, USA: MIT Press, 1999.

[93] A. Astaras, R. Dalzell, A. Murray, and M. Reekie, "Pulse-based circuits and methods for probabilistic neural computation," in *Proceedings of the International Conference on Microelectronics for Neural, Fuzzy and Bio-Inspired Systems*, pp. 96–102, April 1999.

[94] "Bayesian network." url="http://en.wikipedia.org/Bayesian_network", October 2005.

[95] J. Alspector, R. B. Allen, A. Jayakumar, T. Zeppenfeld, and R. Meir, "Relaxation networks for large supervised learning problems," in *Advances in Neural Processing Systems*, vol. 4, pp. 1015–1026, 1991.

[96] H. Chen and A. F. Murray, "A Continuous Restricted Boltzmann Machine with an implementable training algorithm," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 150, no. 3, pp. 153–158, 2003.

[97] G. E. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," *Neural Computation*, vol. 14, pp. 1771–1800, 2002.

[98] H. Chen, P. Fleury, and A. F. Murray, "Minimizing Contrastive Divergence in noisy, mixed-mode VLSI neurons," in *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[99] P. Fleury, H. Chen, and A. F. Murray, "On-chip Contrastive Divergence learning in analogue VLSI," in *Proceedings of the International Joint Conference on Neural Networks*, (Budapest, Hungary), pp. 1723–1728, July 2004.

[100] E. A. Johannessen, L. Wang, L. Cui, T. B. Tang, M. Ahmadian, A. Astaras, S. W. Reid, S. Yam, A. F. Murray, B. W. Flynn, S. P. Beaumont, D. R. S. Cumming, and J. M. Cooper, "Implementation of multichannel sensors for remote biomedical measurements in a microsystems format," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 525–535, March 2004.

[101] B. G. Streetman, *Solid State Electronic Devices*, ch. 5: Junctions, pp. 130–200. Prentice-Hall, 3 ed., 1990.

[102] P. Bergveld, "Development, operation and application of the ion-sensitive field effect transistor as a tool for electrophysiology," *IEEE Transactions on Biomedical Engineering*, vol. 19, pp. 342–351, 1972.

[103] P. Muller, R. Dunn, Y. Mendelson, R. Peura, and E. Sacristan, "A preliminary in-vitro evaluation and comparative study of various tissue pH sensors," in *Proceedings of the Eighteenth IEEE Annual Northeast Bioengineering Conference*, pp. 158–159, March 1992.

[104] L. Bousse, J. Shott, and J. D. Meindl, "A process for the combined fabrication of ion sensors and CMOS circuits," *IEEE Electron Device Letters*, vol. 9, pp. 44–46, January 1988.

[105] J. Bausells, J. Carrabina, A. Errachid, and A. Merlos, "Ion-sensitive field-effect transistors fabricated in a commercial CMOS technology," *Sensors and Actuators B*, vol. 57, pp. 56–62, 1999.

[106] Y. L. Chin, J. C. Chou, T. P. Sun, W. Y. Chung, and S. K. Hsiung, "A novel pH sensitive ISFET with on chip temperature sensing using standard CMOS process," *Sensors and Actuators B*, vol. 76, pp. 582–593, 2001.

[107] C. G. Jakobson, U. Dinnar, M. Feinsod, and Y. Nemirovsky, "Ion-sensitive field-effect transistors in standard CMOS fabricated by post processing," *IEEE Sensors Journal*, vol. 2, pp. 279–287, August 2002.

[108] P. A. Hammond, D. Ali, and D. R. S. Cumming, "Design of a single-chip pH sensor using a conventional $0.6 \mu m$ CMOS process," *IEEE Sensors Journal*, vol. 4, pp. 706–712, December 2004.

[109] DuPont, "Nafion® membranes and dispersions," April 2006. URL:http://www.dupont.com/fuelcells/products/nafion.html.

[110] H. X. Zhu, T. C. Lo, R. Lenigk, and R. Renneberg, "Fabrication of a novel oxygen sensor with CMOS compatible processes," *Sensors and Actuators B*, vol. 46, pp. 155–159, 1998.

[111] M. E. Sandison, N. Anicet, A. Glidle, and J. M. Cooper, "Optimization of the geometry and porosity of microelectrode arrays for sensor design," *Analytical Chemistry*, vol. 74, no. 22, pp. 5717–5725, 2002.

[112] A. Bard and L. Faulkner, *Electrochemical methods: fundamentals and applications*. New York: John Wiley & Sons, 1980.

[113] B. E. Conway, "Double layer capacitor," March 2003. http://electrochem.cwru.edu/ed/dict.htm.

[114] S. M. Martin, F. H. Gebara, T. D. Strong, and R. B. Brown, "A low-voltage chemical sensor interface for system-on-chip: The full differential potentiostat," in *Proceedings of International Symposium on Circuits and Systems*, vol. 4, (Vancouver, Canada), pp. 892–895, May 2004.

[115] R. F. B. Turner, D. J. Harrison, and H. P. Baltes, "A CMOS potentiostat for amperometric chemical sensors," *IEEE Journal of Solid-State Circuits*, vol. 22, pp. 473–478, June 1987.

[116] P. A. Comte and J. Janata, "A field-effect transistor as a solid-state reference electrode," *Analytica Chimica Acta*, vol. 101, pp. 247–252, November 1978.

[117] T. Matsuo and H. Nakajima, "Characteristics of reference electrodes using a polymer gate ISFET," *Sensors and Actuators*, vol. 5, pp. 293–305, July 1984.

[118] D. Wilhelm, H. Voigt, W. Treichel, R. Ferretti, and S. Prasad, "ph sensor based on differential measurements on one pH-FET chip," *Sensors and Actuators B*, vol. 4, no. 1-2, pp. 145–149, 1991.

[119] S. V. Dzyadevych, T. M. Anh, A. P. Soldatkin, N. D. Chien, N. Jaffrezic-Renault, and J. M. Chovelon, "Development of enzyme biosensor based on pH-sensitive field-effect transistors for detection of phenolic compounds," *Biochemistry*, vol. 55, no. 1-2, pp. 79–81, 2002.

[120] W. Olthuis, A. J. Sprenkels, J. G. Bomer, and P. Bergveld, "Planar interdigited electrolyte-conductivity sensors on an insulating substrate covered with Ta_2O_5," *Sensors and Actuators B*, vol. 43, pp. 211–216, 1997.

[121] W. M. Arnold, "Dielectric measurements using novel three-electrode technique," in *Annual Report of Conference on Electrical Insulation and Dielectric Phenomena*, vol. 1, (Atlanta, GA, USA), pp. 356–359, October 1998.

[122] S. N. Nihtianov, G. P. Shterev, N. Petrov, and G. C. M. Meijer, "Impedance measurements with second-order harmonic oscillator for testing food sterility," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, pp. 976–980, August 2001.

[123] M. Ahmadian, B. W. Flynn, A. F. Murray, and D. R. S. Cumming, "Data transmission for implantable microsystems using magnetic coupling," *IEE Proceedings on Communications*, 2004. in press.

[124] A. Astaras, T. B. Tang, L. Wang, S. P. Beaumont, A. F. Murray, and D. R. S. Cumming, "Noise analysis on integrated multisensor microsystems," in *Proceedings of the third IEEE Sensors Conference*, (Vienna, Austria), October 2004.

[125] X. Ouyang and M. G. Amin, "Short-time fourier transform receiver for nonstationary interference excision in direct sequence spread spectrum communications," *IEEE Transactions on Signal Processing*, vol. 49, pp. 851–863, April 2001.

[126] N. Aydin, T. Arslan, and D. R. S. Cumming, "Direct sequence CDMA based wireless interface for an integrated sensor microsystem," in *Proceedings of the IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine*, pp. 370–373, April 2003.

[127] L. Wang, N. Aydin, A. Astaras, M. Ahmadian, P. A. Hammond, T. B. Tang, E. Johannessen, T. Arslan, S. P. Beaumont, B. W. Flynn, A. F. Murray, J. M. Cooper, and D. R. S. Cumming, "A sensor system on chip for wireless microsystems," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2003. under review.

[128] T. B. Tang, E. Johannessen, L. Wang, A. Astaras, M. Ahmadian, A. F. Murray, J. M. Cooper, S. P. Beaumont, B. W. Flynn, and D. R. S. Cumming, "Toward a miniature wireless integrated multisensor microsystem for industrial and biomedical applications," *IEEE Sensors Journal: Special Issue on Integrated Multisensor Systems and Signal Processing*, vol. 2, pp. 628–635, December 2002.

[129] J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, "'Unlearning' has a stabilizing effect in collective memories," *Nature,* vol. 304, pp. 158–159, July 1983.

[130] S. Grossberg, "Competitive learning: From interactive activation to adaptive resonance," *Cognitive Science,* vol. 11, pp. 23–63, 1987.

[131] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," *The Psychology of Learning and Motivation,* vol. 24, pp. 109–165, 1989.

[132] R. Ratcliff, "Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions," *Psychological Review,* vol. 97, pp. 285–308, 1990.

[133] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends in Cognitive Science,* vol. 3, no. 4, pp. 128–135, 1999.

[134] A. Robins, "Catastrophic forgetting, rehearsal and pseudorehearsal," *Connection Science,* vol. 2, pp. 123–146, 1995.

[135] J. McClelland, B. McNaughton, and R. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory," *Psychological Review,* vol. 102, pp. 419–457, July 1995.

[136] T. B. Tang, H. Chen, and A. F. Murray, "Adaptive stochastic classifier for noisy pH-ISFET measurement," in *Proceedings of the 13th International Conference on Artificial Neural Networks,* (Istanbul, Turkey), pp. 638–645, June 2003.

[137] G. E. Hinton, "Products of experts," in *Proceedings of the 9th International Conference on Artificial Neural Networks,* (Edinburgh, Scotland), pp. 1–6, June 1999.

[138] P. Smolensky, *Parallel Distributed Processing: Explorations in Microstructure of Cognition,* vol. 1, ch. Information processing in dynamical systems: Foundations of harmony theory, pp. 195–281. The MIT Press, 1986.

[139] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics,* vol. 22, no. 1, pp. 79–86, 1951.

[140] W. T. Yee and G. E. Hinton, "Rate-coded restricted boltzmann machines for face recognition," in *Advances in Neural Information Processing Systems,* vol. 12, 2000.

[141] G. Mayraz and G. E. Hinton, "Recognizing hand-written digits using hierarchical products of experts," in *Advances in Neural Information Processing Systems,* vol. 13, 2001.

[142] H. Chen and A. F. Murray, "A Continuous Restricted Boltzmann Machine with a hardware-amenable learning algorithm," in *Proceedings of the International Conference on Artificial Neural Networks,* (Madrid, Spain), pp. 358–363, August 2002.

[143] W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–137, 1943.

[144] B. Widrow and M. E. Hoff, "Adaptive switching circuits," *IRE WESCON Convention Record*, pp. 96–104, 1960.

[145] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, pp. 386–408, 1958.

[146] C. E. Rasmussen and Z. Ghahramani, "Occam's razor," in *Advances in Neural Information Processing Systems*, vol. 13, pp. 294–300, 2001.

[147] P. C. D. Fleury and A. F. Murray, "Mixed-Signal VLSI Implementation of the Products of Experts' Contrastive Divergence Learning Scheme," in *Proceedings of the IEEE International Symposium on Circuits And Systems*, (Bangkok, Thailand), pp. 653–656, May 2003.

[148] B. L. Lu, J. Shin, and M. Ichikawa, "Massively parallel classification of single-trial EEG signals using a min-max modular neural network," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 551–558, March 2004.

[149] N. Acir, I. Oztura, M. Kuntalp, B. Baklan, and C. Guzelis, "Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 52, pp. 30–40, January 2005.

[150] D. R. Wilson and T. R. Martinez, "The need for small learning rates on large problems," in *Proceedings of the International Joint Conference on Neural Networks*, vol. 1, (Washington, DC, USA), pp. 115–119, July 2001.

[151] H. C. Demmin, R. B. Mazarella, and J. D. Borkman, "Neural network compensation for sensors." United States Patent, September 1996. US Patent 5554273.

[152] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural network design*. Boston, MA, USA: PWS Publishing, 1996.

[153] M. G. Buehler, G. M. Kuhlman, D. Keymeulen, and S. P. Kounaves, "Advanced electronic tongue concept," in *IEEE Aerospace Conference Proceedings*, vol. 1, (Big Sky, MT, USA), pp. 407–416, March 2002.

[154] S. Holmin, F. Bjorefors, M. Eriksson, C. Krantz-Rulcker, and F. Winquist, "Investigation of electrode materials as sensors in a voltammetric electronic tongue," *Electronanalysis*, vol. 14, pp. 839–847, July 2002.

[155] K. Persaud and G. Dodd, "Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose," *Nature*, vol. 299, pp. 352–355, September 1982.

[156] K. Mitsubayashi and Y. Hashimoto, "Bioelectronics sniffer device for trimethylamine vapor using flavin containing monooxygenase," *IEEE Sensors Journal*, vol. 2, pp. 133–139, June 2002.

[157] G. Bernath, "Fault detection and isolation," *ATR Australian Telecommunication Research*, pp. 35–43, February 1994.

[158] R. Dettmer, "Self validation in process control," *IEE Review*, vol. 46, pp. 29–32, July 2000.

[159] J. Frolik, M. Abdelrahman, and P. Kandasamy, "A confidence-based approach to the self-validation, fusion and reconstruction of quasi-redundant sensor data," *IEEE Transactions on Instrumentation and Measurement*, vol. 50, pp. 1761–1769, December 2001.

[160] S. Amari, "A theory of adaptive pattern classifiers," *IEEE Transactions on Electronic Computers*, vol. 16, no. 3, pp. 299–307, 1967.

[161] N. Murata, K. Muller, A. Ziehe, and S. Amari, "Adaptive on-line learning in changing environments," in *Advance in Neural Information Processing Systems*, vol. 9, pp. 599–605, 1996.

[162] T. B. Tang, H. Chen, and A. F. Murray, "Adaptive, integrated sensor processing to compensate for drift and uncertainty: a stochastic 'neural' approach," *IEE Proceedings on Nanobiotechnology*, vol. 151, no. 1, pp. 28–34, 2004.