CIRCUITS AND SYSTEMS FOR INTELLIGENT HEARING AIDS

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

by

Ian Daniel Hasler

June 2009

ABSTRACT

This thesis presents research and development of electronic circuits and systems that will be used for a new generation of hearing aids. The next generation of hearing aid devices will implement a high level of adaptive signal processing enabling the hearing aids to have excellent performance and flexibility. The thesis brings together research on the natural human auditory system and hearing impairment together with research into developing a hearing aid that uses the human auditory system as a system model for the design of a hearing aid circuits and systems. This thesis then focuses on the implementation of the required signal processing with the criteria of low power; low distortion; flexibility and small scale. The work concluded that a hybrid mixed analogue/digital signal processing approach is most efficient. Hybrid signal processing uses the very low power consumption advantages from each of analogue and digital signal processing to achieve a system which operates in a very similar way to the highly power efficient human biological signal processing systems. The use of a hybrid mixed signal approach leads to the development of silicon on insulator (SOI) technology to implement hearing aid circuits because of its performance with a mixed signal system and possibilities for silicon micro-machined (MEM) microphones to provide a complete system on chip. The thesis identifies the well known Lyon & Mead, cochlea model as a highly effective and power efficient hybrid frequency feature extraction hearing aid system that uses a cascade of analogue filters. The thesis then presents, in depth, SOI as the technology for these analogue filters. The work focuses on modelling analogue circuits in SOI and operating SOI devices in weak inversion. The thesis highlights the importance of the dynamic range and its dependence on noise and linearity of the component operational transconductance amplifiers (OTA). There is therefore an identification of the noise performance of SOI and a SOI OTA with a 10 fold improvement in linear input range is presented. The thesis presents a comparison of fully–depleted SOI versus traditional Si bulk CMOS for implementing analogue hearing aid filters. Finally in this thesis a DSP adaptive filter hearing aid noise cancellation system is investigated and the results show good noise cancellation for a wide range of SNRs - whilst the DSP can be simply and power efficiently implemented.

CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

The author wishes to thank Doctor Octavian Buiu for his supervision and guidance both academically and personally. The author would like to thank Professor S. Hall for this advice and comments to this thesis. The author would like to acknowledge his gratitude to the members of UCL, Microelectronics Laboratory, Louvain-la-Neuve, Belgium for their expert help and assistance, in particular the author could like to acknowledge the help of Bertrand Rue and Professer D. Flandre. The author also wishes to acknowledge ARM UK, EUROSOI and EPSRC for their financial assistance.

This thesis is dedicated to my wife, parents and Christopher, Karl and Samuel for their endless support and encouragement. They are very special people.

PUBLICATIONS

I. Hasler, O. Buiu, "Design of a noise cancellation system for intelligent hearing aids", Proc. of Int. CAS 2004 Proceedings, pp. 479-482, no. 2, 4-6 Oct. 2004, Sinaia, Romania.

I. Hasler, O. Buiu, B. Rue, D. Flandre, S. Hall, "Design of a low voltage – low power highly linear fully depleted SOI OTA for hearing aid applications", Proc. of EUROSOI 2006, pp. 27-28, March 2006, Grenoble.

I. Hasler, O. Buiu, B. Rue, D. Flandre, S. Hall, "Noise in Fully - depleted and Linearized SOI OTAs based G_m-C filters", Proc. of EUROSOI 2007, pp.76-77, Jan. 2007, Louvain.

GLOSSARY

Δf	bandwidth
ΔI	differential current
1/f	flicker noise
ABI	auditory brainstem implant
ACE	advanced combination encoder cochlea implant coding strategy
ACM	device simulation model
ADC	analogue to digital converter
AF	process empirical parameter
AGC	automatic gain control
ANC	adaptive noise canceller
ASIC	application specific integrated circuit
A_{vo}	open loop gain
B	transconductance bandwidth proportionality constant
BAHA	bone anchored hearing aid
BOX	buried oxide layer
BTE	behind the ear hearing aid fitting
c	distance along the channel
C	capacitor
C'_{ox}	device gate oxide capacitance per unit area
CA	compressed analogue cochlear implant coding strategy
Ca^{+}	calcium ion
C_{dbj}	device 'j' drain bulk capacitance
C_{dep}	device depletion layer capacitance
C_{depsoi}	device depletion layer capacitance of the FD SOI layer
C_{gdj}	device 'j' gate drain capacitance
C_{gsj}	device 'j' gate source capacitance
CI	cochlea implant
CIC	complete in canal hearing aid fitting
CIS	continuous interleaved sampling cochlea implant coding strategy
C_L	load capacitance
C-level	maximum power threshold for hearing loud sounds at a particular frequency
CMOS	Complementary metal–oxide–semiconductor
C_{ox}	device gate oxide capacitance
C_{oxb}	device buried oxide layer capacitance
$d(i)$	desired signal
DAC	digital to analogue converter
D_e	diffusion coefficient for electrons
$DIBL$	drain-induced barrier lowering

DSP	digital signal processing
DRC	dynamic range compression
$e(i)$	error signal
$E\{j\}$	Expected signal j
EEPROM	electrically erasable programmable read-only memory
e_{eq}	equivalent input noise potential
EKV	EPFL device simulation model
e_{nj}	device 'j' equivalent noise
e_{th}	thermal noise potential
f	frequency
F1, F2, F3	first, second, third format of speech
FD SOI	fully depleted silicon on insulator wafer technology
f_{gm}	transconductance bandwidth frequency
FIR	finite impulse response
FPGA	field programmable gate array
f_t	transition frequency
g'_m	effective transconductance
GBW	gain bandwidth product
g_m	device transconductance
G_m-C	transconductance–capacitor filter
g_{mj}	device 'j' transconductance
g_{mOTA}	OTA transconductance
g_Q	transconductance of OTA that controls quality in filter
g_t	transconductance of OTA that controls transition frequency in filter
h	transconductance proportionality constant
$i_{1/f}$	flicker noise current
I_b	OTA bias current
I_{bQ}	bias current to control of OTA that controls quality in filter
IC	inversion coefficient
i_{cnf}	carrier number fluctuation theory noise
I_D	device drain current
I_{dj}	device 'j' drain current
IHC	inner hair cells
I_{in}	input current
I_n	noise current
i_{mf}	mobility fluctuation theory noise current
I_o	device process dependant constant for weak inversion
i_{out}	output current
i_{sn}	shot noise current
ITC	in the canal hearing aid fitting
ITE	in the ear hearing aid fitting
i_{th}	thermal noise current
i_{to}	total output noise current
k	Boltzman constant

K	supply voltage reduction constant
K^+	potassium ion
K_F	empirical constant
KF	process empirical parameter
l	filter length
L	device length dimension
LMS	least mean squared algorithm for adaptive filters
$LVLP$	low voltage, low power
m	width-to-length ratio M1 & M2
MEI	middle ear implant
MEM	micro electro mechanical system
M_j	device 'j'
MPS	hybrid of CA and CIS cochlea coding strategy
MSE	mean squared error
n	body effect coefficient
Na^+	sodium ion
N_a	substrate doping
n_e	effective body effect coefficient
NLMS	normalised least mean squared algorithm for adaptive filters
n_o	body effect coefficient for the technology
N_{ot}	effective number of traps
$N_T(E_{fn})$	number of taps at quasi – Fermi level
OHC	outer hair cells
OTA	operational transconductance amplifier
P	power
PD SOI	partial depleted silicon on insulator
p_j	OTA pole 'j'
PNPN	P-type to N-type to P-type to N-type semiconductors junctions
PPS	hybrid of CA and CIS cochlea coding strategy
q	electron charge
Q	filter quality factor
Q_B	device depletion layer charge
Q_I	device inversion layer charge
Q_{I0}	device inversion charge at the source
Q_{IL}	device inversion charge at the drain
RISC	reduced instruction set computer
RLS	recursive least square algorithm for adaptive filters
rms	root mean squared
r_o	device output resistance
R_{out}	OTA output resistance
R_{SD}	source degradation device resistance
R_{th}	equivalent resistor in which the thermal noise occurs
R_{xx}	autocorrelation matrix
s.i.	strong inversion

SD	source degradation
SF	NLMS stability factor
Si bulk	silicon bulk wafer technology
SIMOX	separation by implanted oxygen FD SOI wafer production process
SNR	signal to noise ratio
SoC	system on chip
SOI	silicon on insulator wafer technology
SPEAK	example of a maximum spectral peak extraction cochlea implant coding strategy
SRT	speech reception threshold
T	temperature
THD	total harmonic distortion
t_{ox}	gate oxide thickness
T-level	minimum power threshold for hearing quiet sounds at a particular frequency
UNIBOND	FD SOI wafer production process
U_T	thermal voltage
vd_e	volume density of electrons
vd_{eo}	equilibrium number of electrons in the substrate
V_A	device early voltage
V_B	device bulk voltage
V_b	OTA bias current setup voltage
V_{dd}	positive supply voltage
V_{DS}	device drain – source voltage
V_{dssat}	device drain source saturation voltage
V_G	device gate voltage
V_{GS}	device gate – source voltage
V_{GSj}	device 'j' gate – source voltage
V_{ic}	common input voltage
V_{id}	differential input voltage
VLSI	very large scale integration
V_{out}	output voltage
V_{pp}	peak to peak voltage
V_{sat}	saturation or pinch-off voltage
V_{ss}	negative supply voltage
V_{TO}	device threshold voltage
W	device width dimension
w.i.	weak inversion
w_i	filter weights
$x(i)$	input signal
$y(i)$	output signal
z	OTA zero
Z^1	delayed process
α'	scattering coefficient

α_H	Hooge's constant
α_t	$g_Q/(2g_t)$
β	$\mu C_{ox}W/L$
Γ_{rch}	effective channel resistance coefficient
γ'	attenuation coefficient of electron waves into oxide
ε_o	permittivity of free space
ε_{ox}	permittivity of oxide layer
ε_r	relative permittivity
ε_s	permittivity of silicon layer
λ_{max}	maximum eigenvalue
μ	device effective mobility
μ_c	device surface mobility
μ_{eff}	effective mobility
μ_l	mobility if only lattice scattering were present
μ_o	mobility
μ_{ssp}	set size parameter
$\xi(i)$	mean squared error
ξ_{ox}	electric field in oxide layer
ξ_s	electric field in silicon layer
τ	time constant
ψ_s	device surface potential
ω_o	cut-off frequency
∇	gradient operator
$\cdot^*$	transposed matrix operator
$\cdot^H$	complex conjugate transposed matrix operator

Chapter 1

1 INTRODUCTION

1.1 Motivation

The starting point for this thesis is the natural human auditory system. We must study the natural human auditory system as a biological engineered system so that we can develop hearing aids that can compliment the biological system with an electronic auditory system. The electronic auditory system or hearing aid can therefore be used to overcome an individual's biological auditory system's deficiencies. Unfortunately statistics show only approximately 25% of people who could benefit from a hearing aid have one and of these as few as 53% are satisfied with their hearing instrument. Therefore there is plenty of potential to provide more satisfactory and more effective hearing aid circuits and systems. The ideal specification for hearing aid technology requires them to be unobtrusive (small scale), power efficient and easy to use. Hearings aids must also be highly adaptive systems to operate effectively with different types of hearing loss and in all types of noise environments. The motivation for this thesis is therefore to develop circuits and systems for hearing aids that can most effectively use technology to meet this specification.

1.2 Research Aims

The objective of this research is to develop hearing aid circuits and systems to best implement an electronic auditory system to overcome an individual's auditory deficiencies. The approach used in this research will enable innovation in the development of circuits and systems for hearing aids and their implementation in hardware.

The thesis will take the human auditory system as an inspiration and model for developing a hearing aid. The research will identify how the natural human auditory system operates as a system and studies current auditory

models and hearing aid technologies. This study focuses on assessing their characteristics and limitations. This will enable the fundamental requirements for effective hearing aid circuits and systems to be identified. This knowledge will them be used to develop technology for implementing these circuits and systems within hearing aids.

1.3 Contribution of this Work

The project involved researching the electronic circuits and systems that would be required for a new generation of hearing aids.

This thesis shows how the natural human auditory system can be used as a model and basis for developing hearing aid circuits and systems. It shows how previously developed electronic auditory models can be integrated into the core of a hearing aid device. The research identifies the specification for the implementation of hearing aid circuits and systems in terms of efficient power consumption, price, performance and flexibility. The specification provides no clear choice in terms of analogue versus digital implementations. This leads to the conclusion that a hybrid mixed signal implementation would provide the best solution for this application. The mixed signal approach means specific attention is required in terms of the technology platform for the hardware implementation, i.e. Silicon on insulator, SOI, technology. SOI technology provides good integration of the DSP and analogue circuits with MEM microphones or arrays of microphones, to produce a truly complete system on chip, SoC, solution.

The thesis shows how fully depleted, FD, SOI technology performs when it is used to implement analogue filter circuits as part of a hybrid system for frequency feature extraction within a hearing aid. The traditional view of SOI technology for this kind of analogue circuit would conclude that there is a detrimental effect on performance compared to a traditional silicon bulk, Si bulk, technology [1.1]. This reduction in performance results from the reduction in transconductance associated with SOI transistor structures. This thesis however shows that it is important to study other factors which enhance the performance of SOI circuit

compared to Si bulk. As a result overall it is possible to achieve comparable performance for an SOI and Si bulk circuit.

The thesis also identifies how adaptive filters can be used with a hearing aid system to remove noise and promote the recognition and understanding of speech. The thesis develops an adaptive noise cancellation system for a hearing aid application.

1.4 Organisation of Thesis

The thesis begins by looking at the natural human auditory system studying how it functions and operates as a biologically engineered system. As part of the investigation into the human auditory system it is necessary to have an understanding of characteristics of the main input signals, i.e. speech and noise. The natural human auditory system is translated into a system model of various functional blocks. This model is then used as the inspiration for a hearing aid systems model. The model can then be used to create an electronic device which can overcome deficiencies and malfunctions in the natural auditory system whilst still being able to fit and interact with the existing natural auditory system.

Following the creation of the hearing aid system model the thesis discusses the circuits and systems required to implement stages within the model. Taking the fact that the overall model is based on the natural auditory system the thesis looks at previous work on creating electronic versions of the human ear. An electronic auditory model can then be adapted and developed to form part of a hearing aid.

Based on the individual's requirement and current hearing aid technology a specification for implementing circuits and systems within a hearing aid is discussed. From this it becomes clear there is no clear justification for using either exclusively analogue or exclusively digital circuits and a biologically inspired hybrid approach is most efficient. Therefore a technology platform for a hybrid mixed signal approach is required. A

SOI technology platform can offer performance advantages for these kinds of mixed signal circuits and systems.

Less is known about the performance of SOI technology for analogue circuits than for digital circuits. Therefore the thesis specifically compares the implementation of transconductance–capacitance, G_mC, analogue filter circuits in FD, SOI MOSFET with traditional Si, bulk MOSFETS. These filter circuits form part of a "frequency feature extraction" stage within the hearing aid model. The performance is specifically compared in terms of dynamic input range, noise tolerance, and supply voltage.

This thesis also looks at adaptive filtering for a noise cancellation system within the hearing aid model. An investigation of how adaptive filters operate with this kind of application is presented. This includes a discussion of the algorithm used to adapt the filters within the system. An adaptive noise cancellation system is developed for the hearing aid application.

References

[1.1] Flandre, D., Vanhoenacker, D., "Thin-film fully-depleted SOI CMOS technology, devices and circuits for LVLP analogue/digital/microwave applications", Proc. of Semiconductor Conf., CAS '98, vol. 1, pp. 115-124, 1998.

Chapter 2

2 THE HUMAN AUDITORY SYSTEM

This chapter presents an anatomical and physiological overview of the human auditory system. Hearing is discussed from an analytical point of view highlighting, in particular, those features and characteristics which are relevant to the functioning of hearing aids. This discussion is important for the following chapters. This is because the natural auditory system will provide the inspiration for the design of hearing aid circuits and systems. Also the discussion enables the understanding of when and where hearing impairment manifests within the auditory system and how and where a hearing aid must interface with the natural human auditory system. This chapter discusses and defines speech and noise from a processing viewpoint. This is important because speech and noise are the inputs to the auditory system.

2.1 Human Auditory System

The starting point for considering the hearing aid circuits and systems is the normal functional and operation of the human auditory system. Firstly there must be an understanding of exactly where and within which context a hearing aid must operate and interact with the wearer. Secondly the normally functioning human auditory system is used as a model for an artificial electronic auditory system which will interface with the natural auditory system and therefore operate as a hearing aid. There are obvious reasons for choosing the human auditory system as a template for the hearing aid system, since the hearing aid will be replacing or working alongside, as well as directly interacting with, a human auditory system. However there are also good reasons from an engineering point of view for using the human auditory system as a template for a hearing aid system. The human auditory system has evolved over millions of years to be an

excellent signal processing system in terms of power consumption, sensitivity, dynamic range and noise tolerance.

Normal human hearing is a very sensitive, active and a highly nonlinear processing system which is not yet fully understood. Humans can hear sounds ranging in frequency from approximately 10 Hz to 20 kHz. At the same time, the hearing system can operate as an excellent frequency analyser, differentiating tones whose frequencies vary by less than one percent [2.1]. We can tolerate a wide dynamic range with a 12 times order of magnitude from the faintest audible whisper to the loudest comfortable listening level, and within this range we can discriminate approximately 100 loudness levels [2.2].

Human hearing is primarily a system for the recognition of human speech, to historically aid communication, social understanding and aid learning. We are also able to understand and filter speech from a wide range of sources and in a range of different noise environments. These environments could range from road noise to the babble of a large number of people talking together, for example, in a bar or restaurant. A secondary function of the human auditory system involves the location of the source of sound. We can accurately discern the direction and position of a variety of acoustic sources in various environments [2.3].

2.2 Anatomy and Physiology of the Auditory System

Physiologically the ear is basically a system which translates acoustic energy into electrical nerve energy which is then passed to the brain. Figure 2.1 shows that the anatomy of the ear is split into three parts; the outer; middle and inner ear. The outer ear is comprised of the ear flap made up of the pinna and concha, and the ear canal. It has the function of focusing the acoustic pressure wave energy towards the ear drum. The middle ear contains the ear drum, or tympanic membrane, and three small bones called the Incus, Malleus and Stapes; these are collectively called the ossicles and they are more commonly known as the hammer, anvil and stirrup because of their appearance. The middle ear transforms the

acoustic energy into mechanical vibrations within the inner ear. The inner ear contains the fluid filled snail shaped cochlea and the auditory or cochlea nerve. The basic function of the inner ear is to convert the mechanical energy into processed electrical nerve energy for communication with the central nervous system and transmission of information to the brain. Whilst performing this energy translation the inner ear carries out a number of more complex functions.

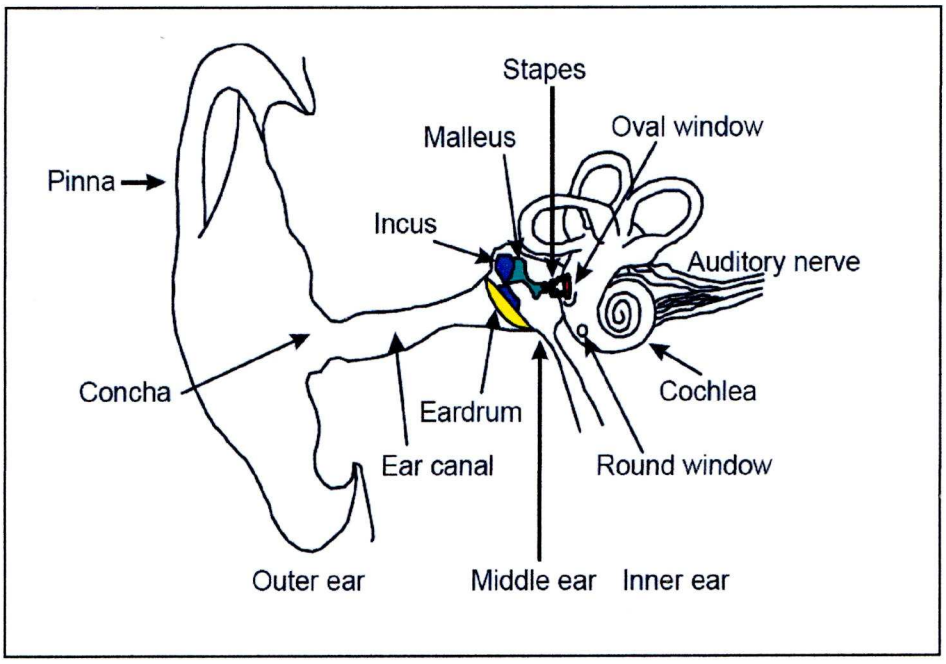

Figure 2.1, The Human Ear (taken from [2.4])

As described later a considerable amount of research has been carried out to understand and model the sophisticated functionality of the cochlea. Despite this research there is still not a complete understanding of how the auditory system within the cochlea precisely operates.

The basic function of the cochlear is firstly as a frequency analyser and feature exactor so that different auditory nerves are stimulated at varying stimulation rates by the frequency components and amplitude features of the acoustic input. The cochlea, and in particular the outer hair cells, OHC, are also believed to have active characteristics. It can control the gain of certain frequency bands resulting in an ability to mask or amplify certain extracted features. This can help mask some speech from noise and

operate as an automatic gain control. Secker-Walker and Searle [2.5] presented this conclusion by showing the perception of vowel forming format frequencies is enhanced by the non-linear filtering of the inner ear.

Figures 2.2 and 2.3 show the basilar membrane which runs the length of the cochlea. This divides the cochlea into two canals, the vestibular and tympanic canals. The membrane is a flexible membrane which is displaced by pressure variations in the fluid within the cochlea [2.6]. The change in width and hence the change in the stiffness of the membrane along its length causes it to act as a frequency analyser. (The membrane is stiff at the base and floppy at the apex). As a result, low frequency energy propagates a long way through the cochlear and causes a peak disturbance of the basilar membrane towards the apex of the cochlea. However high frequency energy only propagates a short way along the cochlear and peaks near the base of the cochlear. In all cases, once the energy has caused a peak disturbance in the basilar membrane, the displacement is quickly damped due to the mechanical properties of the membrane and does not travel further along the membrane. This removes these frequency components from future signal analysis along the membrane, resulting in the basilar membrane being frequency selective in the position at which maximum displacement occurs followed by no displacement further along the membrane.

Situated upon the basilar membrane is the Organ of Corti. This contains two sets of hair cells: the inner hairs cells, IHC, and the outer hair cells, OHC. Taking a cross-section of the Organ of Corti, as shown in figure 2.3, there are three rows of OHC and one row of IHC. The IHC are primary elements which cause electrical stimulation of the auditory nerve fibres with 95% of the fibres interacting with IHC and only 5% interacting with the OHC. Each IHC interacts with approximately 20 auditory nerve fibres. The hair cells possess a tapered bundle of specialised tiny finger-like projections from the surface of the cells called stereocilla. The tips of the stereocilla are mechanically linked by tip elements which are thought, under changing tension, to act as mechanical gate controls which allow

ions to pass into the hair cells via channels in the stereocilla. The stereocilla are also in intimate contact with the tectorial membrane so that displacement of the basilar membrane causes a shear force to be set up between the tectorial membranes resulting in the bending of the bundle of stereocillia. The endolymph fluid in the cochlear surrounding the hair cell tips has a high concentration of potassium K^+ ions and hence has a potential of approximately 90 mV. Therefore as the bundle is bent in one direction the transduction channel gates are opened allowing the influx of K^+ ions into the hair cells causing the cell to become depolarized. This initial depolarization by K^+ ions, allows other channels to open in the body of the cells which allow further depolarization as calcium Ca^{2+} ions enter the cell.

At this stage the major difference between the IHC and OHC will be presented. Studying the IHC first, depolarization causes a channel at the base of the cell to open which releases the transmitter substance called Glutamate from the IHC. This substance is collected in the synaptic cleft between the base of the cells and the end of the auditory nerve fibres. The transmitter substance causes sodium ions Na^+ channels to open in the fibre ends and hence the Na^+ rich Perilymph fluid surrounding the nerve fibre ends enters and potentials within the fibres are created. This stimulation is then sent to the brain via the 30,000 fibres of the auditory nerve [2.6]. It should also be noted that bending of the stereocillia in the opposite direction causes the K^+ channels to be firmly closed and the IHC become hyperpolarized resulting in a reduction of release of transmitter fluid and a reduction in the stimulation of the auditory nerve fibres. The change of polarization within OHC has a very different effect. The outer hair cell exhibits "motility" or movement.

The polarization causes the shape of the OHC to change, either lengthening or shortening as a result of changes in the internal pressure of the cell. This motility of the OHC creates an active component of the auditory system and adds or subtracts from the mechanical energy passed to the inner hair cell. Thus the auditory system can be viewed as having a

gain control system, through the motility of the OHC and its ability to add or subtract from the polarization which occurs in the IHC.

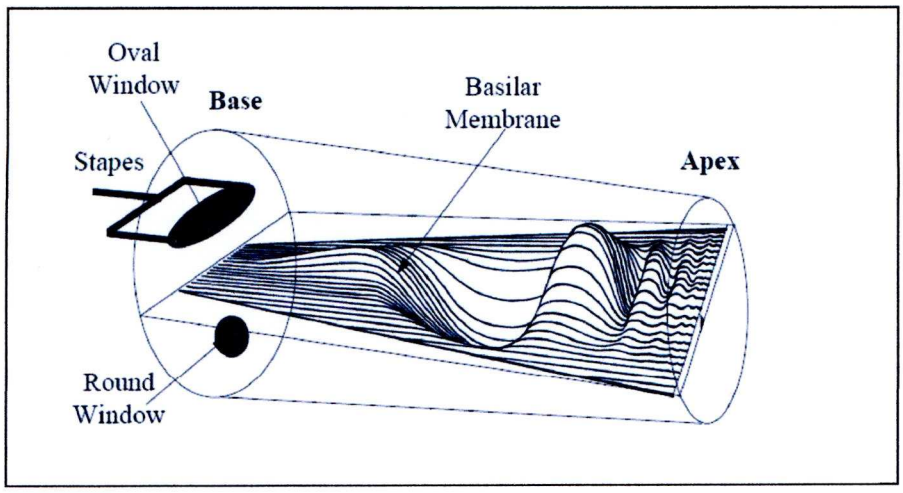

Figure 2.2, Three-dimensional representation of the travelling wave on the basilar membrane in the uncoiled cochlea, (taken from [2.4]).

We currently only have a limited understanding of the complete physiological system of hearing within the brain. However the processing area of the brain responsible for hearing, the auditory cortex, has been identified. It is known that information from the ears goes to both sides of the brain. Physiologically the auditory system is specifically designed for adaptive filtering of speech from noise. As previously discussed it is also believed that the physiology of the inner ear has some functionality for speech perception in noise. The fact we have two ears, one on each side of the head, is also important. Bi-aural hearing is also used to reduce noise by using the different noise references from each ear to filter out the background noise. Bi-aural hearing enables the physiology of the auditory system to localize sound and noise by analysing the time delay between each of the two ears picking up a sound. Localization and noise reduction is achieved by interpreting the differences in the attenuation of the sound signal resulting from each ear caused by sound shadowing of the head [2.7].

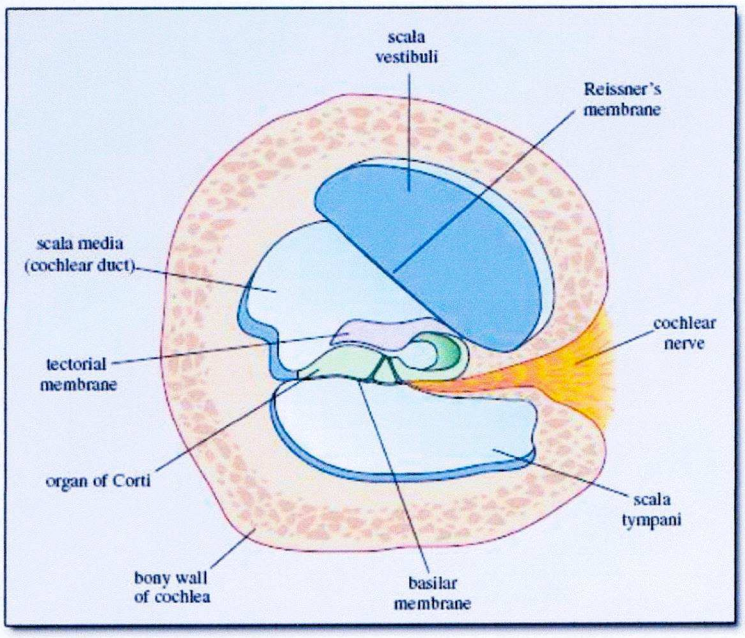

Figure 2.3, Cross section of the uncoiled cochlea, showing the three tubes and the hearing organ, the organ of Corti. Hair cells sit on the basilar membrane and the auditory nerve, (taken from [2.8]).

2.3 Human Speech

The fundamental aim of human hearing is to enable the understanding human speech to enhance and aid communication. Therefore it is important when discussing the human auditory system to have an understanding of speech from a signal processing point of view.

Speech is broken down into a group of sounds associated with each language. These are called phonemes. Each phoneme is created by the repetitive opening and closing of the vocal cords along with differing positioning of the tongue and lips. Voicing is the name given to a set of fundamental frequencies or formants that are a relatively long-lasting phenomenon in speech. These formants vary from person to person and depend on age and sex. Within a particular phoneme there is a signature combination of large concentrations of energy at particular formats. During voicing the spectral or frequency characteristics of a formant change and evolve as phonemes superseded one another. However the frequency of the formants remains relatively unchanged over time. The strong stable formants are easy to distinguish in a spectral analysis. These

are the first three formants called F1, F2 and F3. The fact that these formants remain relatively constant means that from a signal processing point of view, once they have been identified, they can be used to define speech and also define a particular speaker.

2.4 Noise

In almost every real life situation an auditory system, be it the normal human auditory system or a hearing aid, will not just have speech on its own as an input signal. If we define the speech that is to processed and ultimately understood as the "target" signal, then we can define any other input signal as noise. From this definition of noise it can be noted that noise could be other speech signals from people speaking at the same time as the target speaker. When this type of noise from other speaker's effects someone's ability to understand a target speaker it is commonly known as the "cocktail party effect".

As well as other people speaking, noise can be classified into three main categories; "broad-band", "narrow-band" and "transients".

The broad-band noise consists of random noise with energy spread across a wide range of frequencies and is in general representative of the higher frequencies, for example noise experienced within a car, road noise and general noise.

Narrow-band noise comes from the various structural resonances, for example those heard in a car from the engine and gearbox. The spectral analysis is dominated by particularly well defined peaks associated with the mechanical natural frequencies. They are easily identified as not being formant frequency peaks because of their long term, almost stationary, highly focused nature.

Transient noise covers more intermittent noise. It does have a particular regularity, for example car noise due to running over road markings, direction indicator noise and windscreen wiper transients. This noise has a

very regular spectral analysis as it changes in time which can help identify it as not being speech.

It is rare that in any real situation, noise would consist of a single type. In general auditory environments, many types of noise occur simultaneously. These combine to cause a broad-band noise signal where, overall, the noise frequencies are slow changing. This is as opposed to speech signals which have more focused, but still changing frequency components. Also speech can be defined from noise as having overall fast changing frequencies as different phonemes are formed.

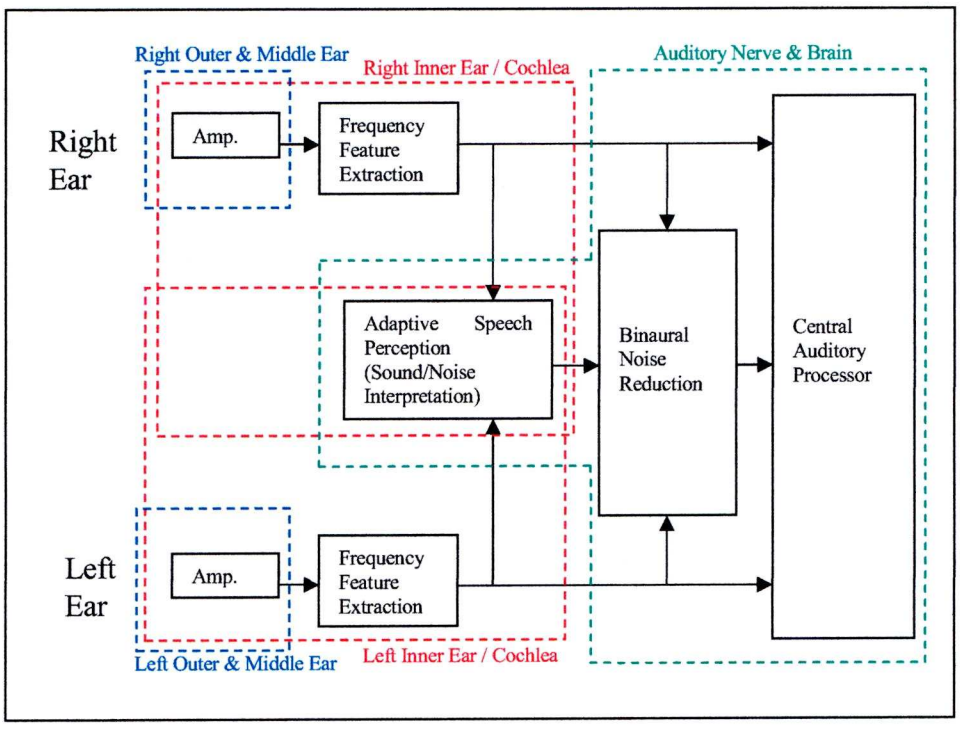

Figure 2.4, Auditory System Process Model.

2.5 Conclusion

The neuro-mechanical process of the human ear is not fully understood. The mechanics of sound entering the outer ear and being mechanically transferred to the inner ear is fairly straight forward. Within the inner ear the mechanical properties of the basilar membrane and fluid mechanics of the sound wave travelling through the cochlea structure acts to analyse the sound wave into its frequency components. This "frequency feature

extraction" is detected by localised stimulation of hair cells along the cochlea and the hair cells in turn stimulate neurons at their bases.

Figure 2.4 is a model of the human auditory system broken down into a series of processing blocks. The figure identifies the location of each processing block within the auditory system as they have been discussed in this chapter.

The outer and middle structures of each ear act to amplify and focus the sound for the inner ear stage. The inner ear, consisting of the cochlear structure, acts as a frequency feature extraction system. The frequency component information is passed via neurons to the brain. Before the brain concludes the processing and understanding of the information there are various processes of noise interpretation and localization which occur partly in the cochlear and partly in the brain itself. These are adaptive intelligent processes which enhance areas of target input sound whilst suppressing other noise input sound. The bi–aural nature of the human auditory system also acts to locate and localize target sound and thereby reduce non–target noise.

This chapter has particularly identified the importance of the frequency feature extraction within the auditory system. Frequency feature extraction is the analysis of an input signal into its frequency components and the analysis of the magnitude of each of these components. The importance of frequency feature extraction for communication can be seen when speech is analysed. Speech can be summarised as a pattern of changing frequency components which are called formats. The exact frequencies of the formats are particular to each individual; however the relationship between formats and the patterns they form are recognised as speech.

The area of the human auditory system that is less well understood is the ability to actively cope with noise. To begin with, it is known that it is very important for an individual to localise sound [2.3] and hence the auditory system is able to target a particular source and to some extent ignore others. This is partly achieved by the shape of the outer ear to focus

incoming sound and more importantly by the functioning of the brain. With regards to hearing aids this has important implications for the position where the devices should be worn and how the microphones should pickup sound, i.e. close to the normal ear position or within the ear canal. On a second level it is recognised that the human auditory system also has an active mechanical process whereby the outer hair cells, controlled by neurons, act to dampen certain frequency components within sound waves. This occurs along with the process of frequency feature extraction within the cochlear and carried out by the inner hair cells. From a signal processing point of view it would make sense for this active mechanical process, within the auditory system, to be working as a noise reduction process. To explain this further, we can think of it as a feedback process whereby the brain is using prior knowledge to recognise frequency components that constitute noise. These frequency components are then dampened or reduced to allow other non-noise, or target frequency components, to stand out and be more significantly recognised. It is easy to conceive that the human brain at a subconscious level would be able to identify noise using the frequency feature extraction process that occurs within the auditory system. The identification of noise would then be used for a feedback noise reduction process based on the different frequency characteristics and slow changing nature of noise compared to speech or another target.

Now that the processing blocks within the human auditory system have been identified the next chapter can discuss the manifestation and details of hearing impairment that a hearing aid device must be designed to overcome.

References

[2.1] Givelberg, E., Bunn, J., "A Comprehensive Three-Dimensional model of the Cochlea", J. Comput. Phys., vol. 191, pp. 377–391, 2003.

[2.2] Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., Javel, E., "Intensity discrimination as a function of stimulus level with electric stimulation", Journal of the Acoustical Society of America, vol. 100, no. 4, pp. 2393-2414, Oct. 1996

[2.3] Hartmann, W. M., "Localization of sound in rooms", Journal of the Acoustical Society of America, vol. 74, no. 5, pp. 1380-1391, Nov. 1983.

[2.4] Van Schaik, A., "Analogue VLSI building blocks for an Electronic auditory pathway", PhD Thesis, Dept. Electronics, Federal Polytechnic of Lausanne, 1997.

[2.5] Secker-Walker, H. E., Searle, C. L., "Time-domain analysis of auditory-nerve-fiber firing rates", Journal of the Acoustical Society of America, vol. 88, no. 3, pp.1427-1436, Nov. 1983.

[2.6] Loizou, P. C., "Introduction to cochlear implants", IEEE Eng. in Med. and Bio. Mag., vol. 18, pp.32-42, 1999.

[2.7] Krahe, R., Larsen, O. N., Ronacher, B., "Directional hearing is only weakly dependent on the rise time of acoustic stimuli", Journal of the Acoustical Society of America, vol. 107, no. 2, pp.1067-1070, Feb. 2000.

[2.8] http://medic.med.uth.tmc.edu/Lecture/Main/ear.htm

Chapter 3

3 HEARING IMPAIRMENT & HEARING AID TECHNOLOGY

Following the discussion of how the human auditory system operates and functions this chapter looks at the sources and manifestations of hearing impairment. The chapter also discusses how hearing impairment is categorised for an individual and the types of circuits and systems that are required within a hearing aid to overcome the sources of hearing impairment. Following the discussion of hearing impairment we are then able to understand relevance and requirements for hearing aids and similar devices.

3.1 Types of Hearing Impairment

There are two separate types of hearing impairment that can occur; Conductive hearing impairment; and Senorineural hearing impairment.

3.1.1 Conductive hearing impairment

Conductive hearing impairment is caused by a disability in the outer or middle ear that causes acoustic energy not to be properly transferred to the inner ear. This can be a defect in the development of the outer and middle ear which means they have not properly formed, or damage to the ear drum and/or ossicles caused by disease or trauma within the middle ear. Generally conductive hearing impairment causes attenuation of input sound and requires a simple amplification hearing aid. However for more serious conductive hearing damage the outer and middle ear can to be bypassed and a hearing aid can be used as an alternative route for the transfer of energy to the inner ear.

3.1.2 Senorineural hearing impairment

Senorineural hearing impairment is the most common type of hearing impairment and is most commonly caused by prolonged exposure to high

volume levels, bacterial and viral infections, fluid build-up, sudden trauma to the ear, as well as through the normal aging process. This type of hearing impairment manifests in the inner ear. There are different ways this impairment occurs.

Impaired inner hair cells can cause loss of sensitivity to acoustic stimulation resulting in a reduction of the acoustic energy level which is passed to the brain. Impairment of the inner hair cells can result in a reduction of the dynamic range that can be perceived, as well as a reduction of the sensitivity to variations in frequency and to overall acoustic energy levels. Hence inner hair cell damage will generate both attenuation and distortion. Hair cell damage is very common, for example, in the USA it accounts for most of the 250,000 people with profound deafness, [3.1].

Within the inner ear, the damage or the simple aging of the basilar membrane can alter its mechanical properties making it stiffer and less elastic. This will again reduce the sensitivity of hearing and reduce the frequency sensitivity and selectivity of the ear, at the frequency feature extraction stage.

To overcome these types of senorineural hearing impairment, a hearing aid is used to compensate for the deficiencies of the inner ear and adjust the acoustic signal so that audibility and perception of the sound is greatly increased.

Within the category of senorineural hearing impairment there can be hearing impairment caused by damage to the brain or the auditory nerve connecting the ear and brain. This type of hearing loss can be caused by trauma or a tumour and can cause severe distortion to hearing and an inability to interpret binaural hearing. This type of hearing impairment is associated with damage to the neuron system and its ability to transfer information. Current technology makes it difficult to design hearing aids that could compensate and take over the role of the neuron system.

3.2 Importance of testing hearing impairment

The exact nature of hearing impairment is very particular to the individual, therefore there are methods of characterising and testing an individual's hearing loss. Understanding how hearing loss can be characterised is important so that hearing aids can be designed to have the flexibility when they are fitted to the characterised individual's hearing loss. It would be no use to have a hearing aid that can be fitted to hearing characteristics that cannot be tested.

There are a number of different tests which are used to characterise the different types of hearing loss. The most important test uses an audiogram. An audiogram uses pure tone signals, (with particular frequencies across the audible range) to determine hearing characteristics at each particular frequency. Within the test the intensity or amplitude of each pure tone signal can be varied. From this the audiogram defines, for each frequency, the threshold for hearing quiet sounds, called the T-level and the threshold of maximum comfortable listening of loud sounds, called the C-level. The audiogram can also be used to categorise the perceived intensity of signals at each particular frequency.

3.3 The Perceived Categories of Hearing Impairment

The discussion about hearing impairment can be simplified (and better organised) if it is categorised into four types of hearing loss: Attenuation, Compression, Perceptual, and Bi-aural.

Taking the model of the hearing aid system presented in chapter 2, figure 2.4, we can use this model to indicate the area of the auditory system that is associated with each category of hearing loss, figure 3.1.

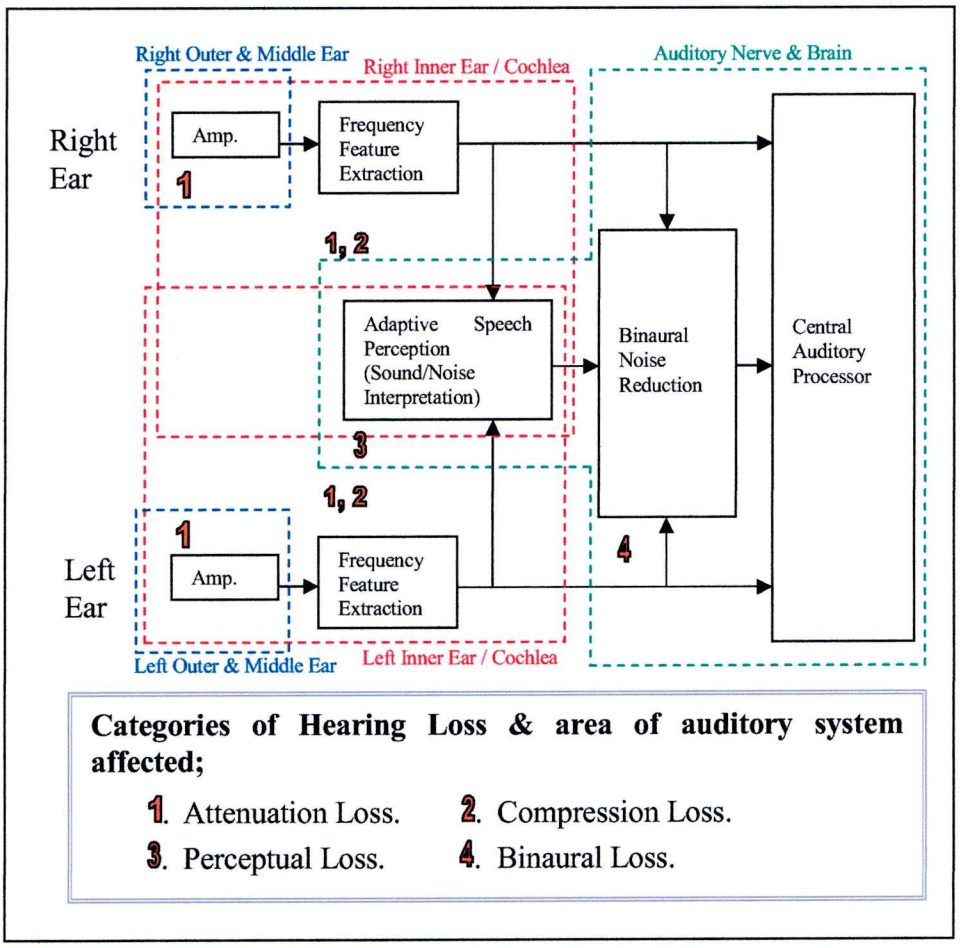

Figure 3.1, Auditory System Process Model.

3.3.1 Attenuation Loss

Attenuation loss can be caused by either conductive impairment within the outer and middle ear or by senorineural impairment within the inner ear. This is when a certain frequency or range of frequencies is attenuated in amplitude resulting in certain sounds and parts of speech being perceived as being quieter.

Attenuation loss is measured using an audiogram. The audiogram determines how well the subject can hear at different frequencies, hence the louder the tone must be in order to be heard, the greater the degree of hearing loss.

To compensate for attenuation loss a hearing aid requires signal processing to amplify the input sound and counteract the attenuation. The simplest

technique is "linear amplification", which means the gain is constant across all sound levels and frequencies. Because of limited dynamic range, linear amplification hearing aids usually include some peak clipping or compression limiting to limit the maximum energy output [3.2]. The disadvantage of the linear amplification is the lack of ability to adapt to different listening environments. Non-linear amplification using Automatic Gain Control, (AGC) improves the ability to interpret soft speech by varying the gain across the frequency range, hence amplifying the weak signals more than the strong ones. This results in the compression of the input dynamic range into a smaller dynamic range at the output [3.2 – 3.4].

3.3.2 Compression Loss

Compression loss is caused by senorineural hearing impairment within the inner ear, where the subject's ear has an inability to hear the full dynamic range of sounds. This results in the subject's threshold of hearing for quieter sounds (T-level) being higher and can also mean their maximum threshold for comfortable listening of loud sounds (C-level) is changed. Compression hearing loss means the subject has a reduced dynamic range of hearing. With a smaller dynamic range, the number of loudness steps that can be differentiated is also reduced.

Using the audiogram setup, compression loss is measured using the so-called 'Categorical Loudness Scaling' which involves the subject being tested with a range of stimuli consisting of different intensity levels spaced equally over the dynamic auditory range. The subjects are then asked to associate each level of intensity to one of a given set of loudness categories. The hearing impairment will be highlighted by a lack of correlation between the signals intensity and the perceived intensity. This test also enables the C-level or threshold of maximum comfortable listening to be defined.

Dynamic Range Compression, (DRC) is the principal hearing aid signal processing that compensates for the loss of range perception by

compressing the full input dynamic range into the limits of the patients dynamic range. Figure 3.2 shows a schematic example of DRC where the input acoustic signal range is translated into a dynamic range that is perceivable by the hearing aid wearer. The upper limit for the compression is the C-level (the maximum intensity for comfortable listening), and the lower limit is the T-level (the threshold of perceivable sound) [3.5]. "Multiband compression" involves separate compression being applied to two or more frequency bands; here the compression is frequency specific and requires the haring aid to have a frequency feature extraction stage. Multiband compression, as opposed to wideband compression, can reduce the masking effect of noise without affecting the audible range of speech at frequencies away from the noise [3.2].

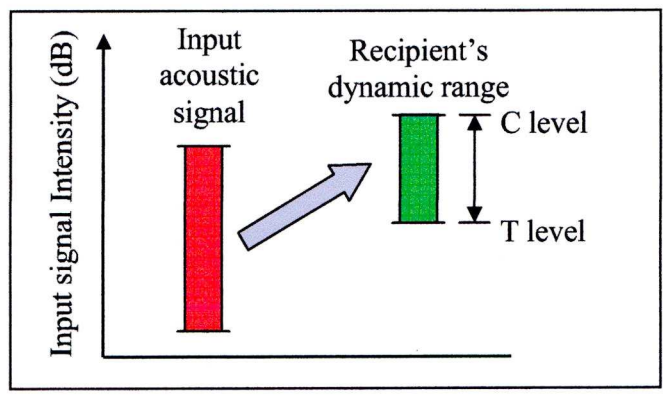

Figure 3.2, Illustration of Dynamic Range Compression.

3.3.3 Perceptual Loss

Perceptual loss is caused by senorineural impairment. Background noise reduces the intelligibility of speech in all cases since the noise will mask fundamental elements of sound that make it intelligible. At high signal to noise ratios, SNR, the noise masking only affects a small amount of the speech and hence the other parts of the speech are sufficient for the auditory system to make the speech intelligible. Conversely as the signal to noise ratio reduces, more of the speech is masked and it becomes less intelligible. Because of the inabilities in the auditory system, a hearing impaired subject suffers a more significant degradation of intelligibility

from noise masking. Hence perceptual loss means subjects have greater difficulty understanding speech in noise.

This category of hearing loss is characterised by measuring a subject's 'Speech Reception Threshold', SRT. In practice this is done by exposing the subject to a set of words which are masked by various signal to noise ratios. The subject is then asked to repeat the word back as it was heard. The result is then a percentage score of the number of correctly perceived words. The words used for this test can take a variety of forms, but the words fall into two types: pure vowel sound (low frequencies), and pure consonant sounds (high frequencies). Single syllable words and sentences can also be tested; also, nonsense-syllables can be used [3.6]. The percentage of correct responses in each case can be used on its own to measure hearing loss. Alternatively, the SRT can be taken as a given percentage point below which it is determined that a subject is unable to perceive the speech at that SNR; this is generally taken to be 50%. The SRT test can also be varied, for example, by using different speakers for the sound, e.g. male voice, female voice or children. Varying speech to noise ratios with varying noise types can be implemented to determine a subject's ability to hear in varying noise environments.

Perceptual loss requires hearing aid circuits and systems that can help distinguish a particular target signal from a noise signal or act to localise and identify a target signal. This requires signal processing that can focus and enhance particular frequencies within the overall input signal. These particular frequencies would belong to the target source whilst other frequencies which contribute to the noise can be suppressed. This kind of signal processing must be adaptive so that it can enhance different targets at different times and in a variety of noise environments. There is also the requirement for the signal processing to identify the target and feedback the knowledge to enable the perception of the target to be enhanced.

The signal processing for perceptual loss often requires adaptive filters as discussed later in this thesis extensively in chapter 8.

3.3.4 Bi-aural Loss

Bi-aural loss is caused by senorineural impairment. This results from impairment in analysing the simultaneous information from both ears. It can again reduce the patient's ability to interpret speech from noise and can cause an inability to localise sound sources.

This is again tested using the SRT test, but this time varying the location of the sound source to determine the directionality of a subject's hearing ability.

Bi-aural loss requires hearing air circuits and systems that can receive sound input from two different sources. The two signals can then be processed to distinguish target sound signals from noise signals or act to enhance the localisation of a particular identified target source. This requires adaptive signal processing which can use a feedback system and prior knowledge to identify a target from the two signals. The input signal can then be processed & filtered to enhance the target and suppress the noise. This kind of two input adaptive filter system usually compares the sum and difference of the two inputs to enable the target and noise to be identified [3.7]. The development of this kind of system is discussed in depth in chapter 8.

3.4 Conclusion

As highlighted in figure 3.1 it can be seen that attenuation loss can manifest in the outer, middle and inner ear. If it manifests from deficiencies in the outer or middle ear a hearing aid to overcome the loss could operate in two ways. Firstly by bypassing the outer or middle ear and passing sound energy (vibrations) straight to the inner ear. Secondly by copying the main function of the outer and middle ear, i.e. act as an amplifier.

It should be noted that most electronic hearing aids will require an amplification stage as part of their signal processing, therefore little effort

is needed to design a hearing aid to compensate for this type of hearing loss as it will be part of any electronic hearing aid system design anyway.

Attenuation, compression and perceptual loss of hearing can all manifest within the inner ear or cochlea. Therefore in order to compensate for these types of hearing loss a hearing aid will require circuits and systems that carry out signal processing which essentially adjusts or replaces the "frequency feature extraction" that is carried out by the cochlea.

Perceptual loss and bi-aural loss would require a hearing aid with electronics which integrates with, and to some extent replaces, the auditory nerve system and brain. This is an obvious area for current and future bio-electronic research [3.8] and beyond the scope of this thesis.

This thesis concentrates on creating hearing aid circuits and systems to overcome the hearing loss that manifests within the inner ear. This is because by designing a hearing aid that can carry out functionality of the inner ear we can use it to compensate for attenuation and senorineural impairment. As previously mentioned, these circuits and systems must carry out signal processing that is equivalent to the "frequency feature extraction" carried out by the natural ear within the cochlear.

The "frequency feature extraction" system processes the input signal into frequency components and analyses the magnitude of each component as identified in the model of the ear presented in figure 2.4, chapter 2. Once an input signal has been processed into frequency components and the features have been identified the hearing aid can adjust and reconstruct the signal to compensate for each kind of hearing loss.

The nature and characteristics of hearing loss within the inner ear are very particular to the individual. Therefore any hearing aid system that involves adjusting the "frequency feature extraction" within the inner ear must to highly flexible and/or adaptive to fit an individual's needs.

References

[3.1] Eddington, D. K., Pierschalla, M. L., "Cochlear Implants, Restoring Hearing to the Deaf", The Harvard Mahony Neuroscience Institute Letter, On The Brain, vol. 3, no. 4, 1994.

[3.2] Lunner, T., Hellgren, J., Arlinger, S., Elberling, C., "Non-linear signal processing in digital hearing aids", Scandinavian Audiology, vol.27, suppl. 49, pp.40-49, 1998.

[3.3] Cochlear Europe Ltd, "Nucleus ESPrit 3G, Whisper Setting", Cochlear Europe Ltd, 22-24 Worpie Road, Wimbeldon, London SW19 4DD, 2002.

[3.4] Scicluna, R., Cornelisse, L. E., "A Programmable DSP Architecture for Hearing Aids", Proc. of the Int. Conf. on Signal Processing Applications and Technology (ICSPAT'99), Orlando, U.S.A.1999.

[3.5] Cochlear Europe Ltd, "Input Signal Processing For Optimal Performance", Cochlear Europe Ltd, 22-24 Worpie Road, Wimbeldon, London SW19 4DD, 2002.

[3.6] Skinner, M. W., Holden, L. K., Holden, T. A., "Speech Recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants", Journal of the Acoustical Society of America, vol. 101, no. 6, pp. 3766-3782, June 1997.

[3.7] Berghe, J. V., Wouster, J., "An adaptive noise canceller for hearing aids using two nearby microphones," Journal of the Acoustical Society of America, vol. 103, no. 6, pp. 3621-3626, June 1998.

[3.8] Schwartz, M.S., Otto, S.R., Brackman, D.E., Hitselberger, W.E., Shannon, R.V., "Use of a multichannel auditory brainstem implant for neurofibromatosis Type 2", Stereotactic and Functional Neurosurgery, vol. 81, pp. 110-114, 2003.

Chapter 4

4 HEARING AID DESIGN AND SPECIFICATION.

This chapter discusses the state of the art and classifications of hearing aid devices. In turn this leads to an analysis of the issues that the next generation of hearing aids must overcome and a specification for the improvement of hearing aid performance.

4.1 Review of Hearing Aid technology

Although, overall, this work will not be based on any previous hearing aid designs or technology, it is obviously logical to review current hearing aid technology and the types and fitting of hearing aids. This will provide a starting point for a specification of size, functionality and cost.

4.1.1 Conventional Hearing Aids

A conventional hearing aid has a microphone input. It carries out signal processing to alter the acoustic signal characteristics of the input acoustic signal and passes this to a small speaker output. The modified acoustic signal is then passed to the ear drum and auditory system as normal. This section will give a brief overview to the categories of conventional hearing aids.

There are four different fitting styles as illustrated in figure 4.1. These different styles of fitting in-or around the ear result in a device being more or less desirable. Desirability generally depends on whether the device is comfortable and unobtrusive. This tends to mean that if the device is smaller it is more desirable. However with miniaturisation the cost increases. The 'Behind the Ear' (BTE) style has a unit worn behind the ear, which houses: the microphone batteries and electronics. The output from the device is generally fed via a tube into a small ear-mould placed in the ear canal. The 'In the Ear' (ITE) style has the unit with microphone, batteries and electronics that all fit inside the ear flap. This unit is

specifically moulded to the wearer's ear and feeds the output directly into the ear canal. The 'In The Canal' (ITC) style has the unit with microphone, batteries and electronics, which this time all fit just inside the entrance to the ear canal. This makes the device far less visible. Lastly the 'Complete In Canal' (CIC) style again has a single unit. However it fits further down the ear canal making it almost invisible. Since the CIC device fits tightly within the canal it is also the least likely of the fitting styles to cause feedback problems. The ear canal for example is approx. 60 mm^3. A less obvious prosthesis can improve self-confidence and avoid prejudice in third parties' attitudes and significantly help children "blend in".

There is also a range of technologies available for conventional hearing aids. As the technology and functionality of the device become more advanced, and consequently more desirable, the cost tends to increase.

The cheapest hearing aids which use simple analogue linear-type amplification technology, cost $395-$795, have the advantage of being low cost and can give benefits (although limited) for patients with a wide range of hearing loss. Non-linear or compression amplification type technology has the ability to limit sound amplification so that it is not uncomfortably loud and it delivers more natural loudness throughout the patients listening range. This technology costs $695-$1295 and these devices have the advantage of improving listening comfort in a range of environments, producing more natural hearing. They also have more additional functions e.g. noise suppression or telephone settings.

Miniaturised low power devices, which are specifically designed for the CIC style, have many of the advantages of non-linear devices with the added benefits of being virtually invisible and very comfortable to wear. They have the capability to have multiple settings which are changed via remote control, require less power as they are positioned close to the ear drum and, because of their fitting style, produce less feedback problems. This type of device costs $1195-$2095.

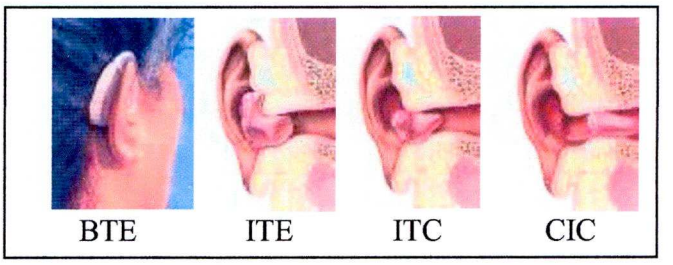

Figure 4.1, Fitting styles for Conventional Hearing Aids, (taken from [4.1]).

Finally there are digital programmable hearing aids which cost $995-$2400 and use digital signal processing to produce flexible and finely tuned listening for patients. They have multiple memories for different listening environments and have automatic and/or self adjustable volume controls. This type of technology also has the ability to be readjusted over time if the patients hearing impairment changes.

4.1.2 Bone Conduction Hearing Aid

Where there is more severe damage to the outer or middle ear a 'Bone Conduction' hearing aid can be used to bypass the outer and middle ear. It is possible to send acoustic vibration to the cochlea by transmitting it through the bone of the skull and jaw. Hence a Bone Conduction hearing aid is attached directly to the bone and converts acoustic energy, picked up from a microphone, and converts the energy to vibrations which are directly transmitted to the inner ear through the bone. There are two types of Bone Conduction hearing aid; the traditional Bone Conduction hearing aid and the Bone Anchored hearing aid, (BAHA). The traditional Bone Conduction hearing aid usually uses a head band which holds a vibration pad in place. The pad must be held tightly in place so these aids are generally uncomfortable and unsightly to wear. They also tend to use a considerable amount of power. A BAHA consists of a small titanium implant which is fixed directly to the bone of the skull just behind the ear. A small hearing aid, approximately 20 mm^3, is then attached directly to this implant. For more severe hearing loss, a body worn amplifier can be used. BAHA's are less visible, use less power and improve audible quality [4.2].

4.1.3 Middle Ear Implants

Another type of hearing aid that can be used for conduction hearing is a middle ear implant, MEI. With an MEI the implant is directly fixed to one of the ossicles in the middle ear. Output vibrations from the hearing aid are then passed directly to the middle ear. The advantage of an MEI is that it keeps the ear canal free. For patients who are sensitive to foreign bodies in their ear canal, for example an ear mould, this is more comfortable. MEI's also have the advantage of eliminating feedback. Conventional hearing aids can create an uncomfortable audible "squeal" resulting from a feedback loop that can occur from the output to the input of the instrument. However additional signal processing using adaptive filtering in conventional hearing aids can be effectively used to remove the problem of feedback [4.3]. It must be noted that the signal processing within conventional hearing aid, BAHA and MEI can be designed to cope with hearing impairment which is part conductive and part sensorineural.

Finally, Park et el. [4.4] describes a proposal for an additional type of conduction hearing aid. This uses a microelectromagnetic vibration transducer as an implantable middle ear hearing aid which amplifies acoustic sound by interacting with the ossicles. Similar to this, Symphonix Devices in San Jose, USA, [4.5] has developed a hearing aid which overcomes feedback problems with a small magnetic implant that is attached to the ossicles. This implant is made to vibrate by a separate ear-mould signal processing unit.

4.1.4 Cochlear Implants

A Cochlea Implant 'CI' uses an implanted array of electrodes that are surgically implanted either inside the cochlea or against the outside of the cochlea to directly stimulate the auditory nerve. These types of hearing aids were first developed and used in the late 1970's by Prof. Graeme Clark at the University of Melbourne Australia and are now used by more than 50,000 children and adults worldwide [4.6]. Generally the electrode array is attached to a data receiver/decoder that is fixed to the skull under the skin. This receiver/decoder collects signals which are transmitted

through the skin from an externally worn sending/encoder and signal processing unit. The signal processing unit which collects the input acoustic signal from a microphone and houses the batteries to run the instrument is either worn behind the ear or as a body worn unit. The alternative is to have the signal processing unit plug directly into the electrode array connection. This has the advantage of significantly reducing the power required by the instrument. However the plug can be extremely difficult to keep free from infection. Various different Cochlear Implants use different numbers of electrodes and different types of coding strategies to stimulate the auditory nerve.

A significant problem with CI devices is that they tend to require a large amount of power and thus require batteries to be changed every 1-2 days. A break through in the cochlear implant technology was the development of cochlear implants such as the "Epic" (developed by EPIC BIOSONICS, Victoria Canada with the help of Prof. Chris Toumazou from Imperial College, London). The Epic is attempting to use low power analogue VLSI processing, equipped with 48 electrodes, which will be totally internal requiring no hard-wired external device or sockets and will use a miniature rechargeable battery which is designed to last a lifetime and will only require 1-2 hrs recharging per week [4.7].

4.1.5 Cochlea Implant coding strategies

An important system within a cochlear implant hearing aid is the coding strategy that is implemented. The coding strategy defines how the characteristics of the input signal are converted to stimulation of the nerves within the cochlea. These strategies and the related systems continue to be developed to enable the hearing aid wearer to better understand the input signal. The coding strategies must analyse and divide the signal into different frequency bands and determine the amplitude relationship within each band. Hence a cochlea implant type hearing aid also requires a "frequency feature extraction system" as defined in structure of a hearing aid presented in this thesis in figure 3.1. The coding strategy must then transform the information from the frequency feature extraction system to

defining where the stimulation occurs, i.e. define the stimulation rate of each electrode and how much to stimulate each electrode to preserve the amplitude relationship. There are four main coding strategies that have been developed; Spectral Maxima Sound Processor, Compressed Analogue, Maximum Spectral Peak Extraction and Continuous Interleaved Sampling [4.8 – 4.10].

Spectral Maxima Sound Processor, (SMSP): This strategy proposes only taking the most important features of speech required for understanding, leaving out other information. In this way an impaired auditory system will not be overburdened. Limiting the amount of information used also reduces the complexity of the processing required. With this strategy, only electrodes associated with a particular feature are stimulated at any one time. This strategy was quite successfully used for a number of years until developments in technology made it possible to process more information rapidly and efficiently, hence allowing more information to be sent and improving listening quality [4.11].

Compressed Analogue, (CA): The major feature of this strategy is that all the electrodes are stimulated simultaneously. The acoustic input is compressed into a smaller dynamic range suitable for electrode stimulation. However all the acoustic information is used. Since all the information is sent simultaneously the strategy requires a lot of power for the processing. Also several problems exist when stimulating electrodes simultaneously because of uncontrolled interaction between closely positioned electrodes. The amplitude mix and frequency overlap are not predictable and can result in considerable listening distortion [4.12].

Maximum Spectral Peak Extraction, (e.g. SPEAK): This is an extended version of feature extraction. The strategy analyses the frequencies and selects a subset of frequency bands, which have the maximum energy. These frequency bands are associated with a different electrode and each time a number of electrodes are selected. In SPEAK, 3 to 10 electrodes are selected and stimulated with an average pulse rate of 250 pulses per second [4.13].

Continuous Interleaved Sampling, (CIS): This is the newest of the strategies. It is a development of the CA strategy in that only one electrode is stimulated at anyone time, but all electrodes are stimulated in sequence with a rapid stimulation rate of more than 850 pulses per second. Hence CIS uses all the acoustic input information and through a rapid stimulation rate provides the auditory nerve and brain with a large amount of information enabling better hearing [4.14].

Other strategies known as MPS or PPS, are a hybrid of CA and CIS together. In MPS, two or more electrodes are stimulated simultaneously but only if they are far enough apart to avoid interaction. Also ACE is a hybrid of CIS and SPEAK. Advanced Combination Encoders (ACE) or n-of-m method uses CIS higher stimulation rate with the 'less than total' electrode selection of SPEAK [4.15].

A considerable amount of research has been carried out to evaluate different strategies and in particular CIS vs. SPEAK [4.16–4.18]. The results of this research have been inconclusive in defining which strategy is most effective for all situations.

4.1.6 Auditory Brainstem Implants

Similar to cochlear implants auditory brainstem implants directly stimulate nerves with electrodes. Research is being carried out on the development of implants which stimulate and send information to the auditory brainstem directly [4.19]. Cochlear Corporation [4.20] has developed the Nucleus® 24 Auditory Brainstem Implant (ABI) which combines the cochlea implant technology with an ABI.

4.2 Developing Hearing Aid technology

We can now use the discussion of hearing aid technology in 4.1 to show how the hearing aid circuits and systems that will be developed in this thesis fit with hearing aid technology.

As discussed in the conclusion to chapter 3 this thesis concentrates on the hearing aid frequency feature extraction system as illustrated in the model for a hearing aid shown in figure 3.1.

The previous discussion shows that a frequency feature extraction system can be used within both a conventional hearing aid with various fitting styles and with a cochlear implant. A frequency feature extraction system for a cochlear implant would be followed by a coding strategy for the stimulation of the nerves via electrodes. The exact nature of this coding strategy is beyond the scope of this thesis, but it can be seen from the study of current coding strategies that the "frequency feature extraction" for a cochlear implant is fundamentally important. Similarly for a conventional hearing aid the "frequency feature extraction" is essentially the hub of the hearing aid signal processing. The signal processing can therefore break down and analyse the input signal and then use adjustment and reconstruction signal processing or a cochlear implant coding strategy.

4.3 Hearing Aid Design Specification

All the work presented in this thesis can now be drawn together to develop a specification for a hearing aid design and in particular for the design of the frequency feature extraction system.

Before looking at a detailed discussion of the specification there is an important note to highlight. This is that although cost is always important in a product that is produced on a large scale, because it is envisaged the next generation of hearing aids will have to be at the forefront of technology, it is expected that they can be high value devices and so cost will not be a primary driving force.

The specification for an ideal hearing aid would;

- have extremely low power consumption to enable long battery life;
- use very low voltage to enable small battery size;

- be highly miniaturised to help with aesthetics; the ear canal for example is approx. 60 mm^3. A less obvious prosthesis can improve self-confidence and avoid prejudice in third parties' attitudes and significantly help children "blend in";
- have vast functionality to provide the best signal processing possible;
- be very flexible to allow good fitting to patients needs in all environments;
- have a high signal to noise ratio, to enable the user to perceive speech easily without noise interference;
- have low distortion of speech to enable the user to communicate and understand speech effectively;
- be low cost.

4.3.1 Challenges for HA development

When thinking about producing a hearing aid design to meet the ideal specification it is found that improving the design in one area often has a detrimental effect on the design in other areas. Figure 4.2 illustrates how all the different criteria of the design influence and affect other areas of the design. For example, as functionality increases so will the power consumption and the product size along with reducing the battery life. Similarly, as the size of the product is reduced the SNR is likely to increase, as well as distortion of the signal, so the power consumption could increase.

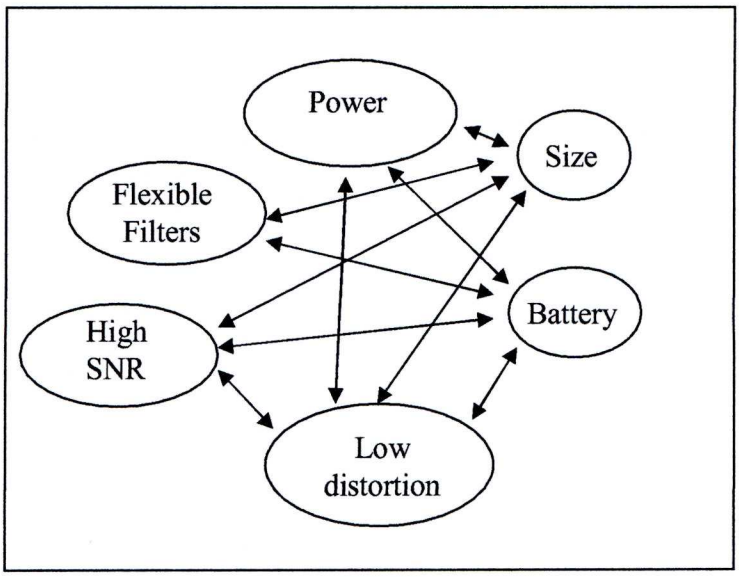

Figure 4.2, Trade off in the design of a Hearing Aid.

4.4 Conclusion

This chapter has discussed the type of hearing aid technology that is available. Following this discussion we can see how the frequency feature extraction system proposed to be developed in this thesis will be the fundamentally important to most of these types of hearing aids.

This chapter develops a specification for hearing aid design and in particular a frequency feature extraction system. The specification identifies a number of trade offs that occur when trying to meet the specification with an ideal design.

The specification listed above can now be used to develop hearing aid circuits and systems. The challenge is to correctly balance all the trade-offs identified in figure 4.2 to get the best possible hearing aid product. This is achieved by gathering together the latest technology for improving the functionality of a hearing aid, and then investigating and demonstrating the best strategy and the best technology to use to implement the functional signal processing.

References

[4.1] http://www.dca.ca.gov/hearingaid/consumer/audiogram.htm

[4.2] RNID Information Services, "Bone conduction hearing aids, Factsheet", Oct. 2000.

[4.3] Hellgren, J., "Analysis of Feedback Cancellation in Hearing Aids With Filtered-X LMS and Direct Method of Closed Loop Identification", IEEE Transaction on Speech and Audio Processing, vol. 10, no. 2, pp. 119-131, Feb. 2002.

[4.4] Park, S., Ki-Chan, L., Byung-Seop, S., Jin-Ho, C., Myoung-Nam, K., Sang-Heun, L., "Microvibration transducer using silicon elastic body for an implantable middle ear hearing aid", Mechatronics, vol. 12, no. 9-10, pp. 1173-84, Nov.-Dec. 2002.

[4.5] Vibrant Soundbridge Implantable Middle Ear Device, VIBRANT MED-EL, Hearing Technology GmbH, Fürstenweg 77, A-6020 Innsbruck, Austria.

[4.6] Eddington, D. K., Pierschalla, M. L., "Cochlear Implants, Restoring Hearing to the Deaf", The Harvard Mahony Neuroscience Institute Letter, On The Brain, vol. 3, no. 4, 1994.

[4.7] Duke, T., "The power of hearing", 2002, www.physicsweb.org

[4.8] Loizou, P. C., "Introduction to cochlear implants", IEEE Eng. in Med. and Bio. Mag., vol. 18, pp. 32-42, 1999.

[4.9] Allum, D. J., "Introduction to Coding Strategies", MED-EL UK Ltd, Bridge Mills, Huddersfield Road, Holmfirth, HD9 3TW, UK.

[4.10] Eddington, D. K., Rabinowitz, W. M., Delhorne, L. A., "Sound processing for cochlear implants", Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, ICASSP, vol. 6, pp. 3449-3452, 2001.

[4.11] McDermott, H. J., Vandali, A. E., Richard, J. M., et al., "Aportable programmable digital sound processor for cochlear implant research", IEEE trans on Reha Eng, vol. 1, pp. 94-100, 1993.

[4.12] Eddington, D. K, "Speech discrimination in deaf subjects with cochlear implants", Journal of the Acoustical Society of America, vol. 68, no. 3, pp. 885-891, 1980.

[4.13] K. B. Nie, J. Liu, and S. K. Gao, "A speech processing strategy for. cochlear implants based on tonal information of Chinese language", IEEE Proc. of Engineering in Medicine and Biology Society, vol. 20, no. 6, pp. 3154-3157, 1998.

[4.14] B. S. Wilson, D. T. Lawson, C. C. Finley, and R. D. Wolford, "Coding strategies for multichannel cochlear implants", Am. J. Oto., vol. 12, suppl. 1, pp. 55-60, 1991.

[4.15] Kaibao, N., Xianhua, H., Ju, L., "Implementation of CIS Speech Signal Processing for Cochlear Implants by Using Bark Scale Frequency Band Partition", in Proc. of ICONIP'01, Shanghai, 2001.

[4.16] Loizou, P., Graham, S., Dickins, J., Dorman, M., Poroy, O., "Comparing the performance of speak strategy (Spectra 22) and the CIS strategy (Med-El) in quiet and in noise", Proc. Conference on Implantable Auditory Prosthese, Asilomar, Monterey, CA., 1997.

[4.17] Young, N. M., Grohne, K. M., Carrasco, V. N., Brown, C., "Speech Perception of Young Children using Nucleus 22-Channel or Clarion Cochlear", Annals of Otology, Rhinology & Laryngology, vol. 108, no. 4.2, pp. 99-103, Apr. 1999.

[4.18] Kiefer, J., Müller, J., Pfenningdorff, T., Schon, F., Helms, J., von Ilberg, C., Baumgartner, W., Gstöttner, W., Ehrenberger, K., Arnold, W., Stephan, K., Thumfart, W., Baur, S., "Speech understanding in quiet and in noise with the CIS speech coding strategy (MED-EL Combi-40) compared to the Multipeak and Spectral Peak strategies (Nucleus)", Journal for Oto-Rhino-Laryngology and Its Related Specialties, vol. 58, pp. 127-135, 1996.

[4.19] Waring, M. D., "Refractory properties of auditory brain-stem responses evoked by electrical stimulation of human cochlear nucleus: evidence of neural generators", Electroencephalography &

Clinical Neurophysiology: Evoked Potentials, vol. 108, no. 4, pp. 331-344, Jul. 1998.

[4.20] Cochlear Corporation,14 Mars Road, PO Box 629, Lane Cove NSW 2066, Australia, http://www.cochlearamericas.com

Chapter 5

5 HEARING AID HARDWARE IMPLEMENTATION

Following the discussion of hearing aid design and the development of a specification for a hearing aid in chapter 4 this chapter discusses the best ways to implement constituent signal processing systems in terms of previous published work and in terms of signal processing hardware technology that is available. The chapter compares the fundamentals of implementing the signal processing computation in analogue VLSI versus digital VLSI hardware.

As previously discussed the thesis will concentrate on hardware implementation of a frequency feature extraction system within a hearing aid as identified in figure 3.1 and discussed in chapter 3. The frequency feature extraction stage can be viewed as the hub of a hearing aid that is designed using the natural human auditory system as a model.

The Lyon & Mead model [5.1] is an electronic model of the human cochlear which in affect achieves the frequency feature extraction stage. As this electronic model is well known and well understood it will be used as a basis for the frequency feature extraction of a hearing aid structure.

5.1 The Lyon & Mead Auditory Model

When designing hearing aid circuits and systems that are based on a model of the natural human auditory system, previous work that has concentrated on modelling the human auditory system electronically can be used

Modelling the auditory system, particularly the cochlea, focuses on the function of the basilar membrane and inner and outer hair cells. As previously discussed, these structures act as a frequency feature extraction system carrying out frequency analysis and separation. The feature extraction modelling must also include the active characteristics which

mask or amplify certain extracted features associated with the outer hair cells [5.2]. The outer hair cells amplify some frequencies and not others as the vibrations propagate along the cochlea. This adjusts some frequencies and limits their energy levels thereby creating automatic gain control and some noise masking with a broad gain peak of as much as 60 dB.

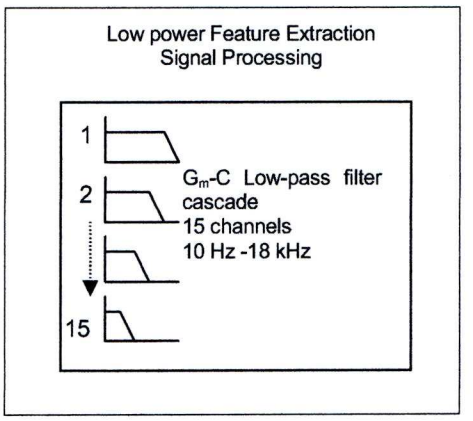

Figure 5.1, Lyon & Mead model of the fluid dynamics of the cochlea as a cascade of filters.

The widely known Lyon & Mead model offers the most appropriate model to be implemented and used effectively as an advanced frequency feature extraction system with controllable channel gain.

This approach models the fluid dynamics of the cochlea as a cascade of filters, figure 5.1. Therefore as the fluid travels along the cochlea and is damped by the basilar membrane characteristics according to the frequency of the signal, this is taken to be the same as a cascade of low pass filters where each channel is used to provide the frequency feature extraction. The effect of the outer hair cells is included as an underdamped filter response for each channel, thereby creating response peak which can be adjusted to control the gain for each channel. This resonant peak is created using second order filter stages with controllable Q–factors enabling gain control of each low pass filter stage. The original work achieved a band-pass gain peak of about 12 dB [5.3]. However, more recent work has improved the gain of the filters to give peak gain close to 40 dB [5.3]. It has also been identified that this peak gain is related to the dynamic range of the filter implementation. A typical feature extraction

stage would have 15 channels with outputs arranged linearly on a log scale ranging from 10 Hz to 18 kHz.

5.2 Signal Processing Hardware Alternatives

There are a variety of signal processing hardware approaches that could be used to implement hearing aid circuits and systems. Figure 5.2 shows a simple breakdown of the general signal processing approaches that can initially be considered.

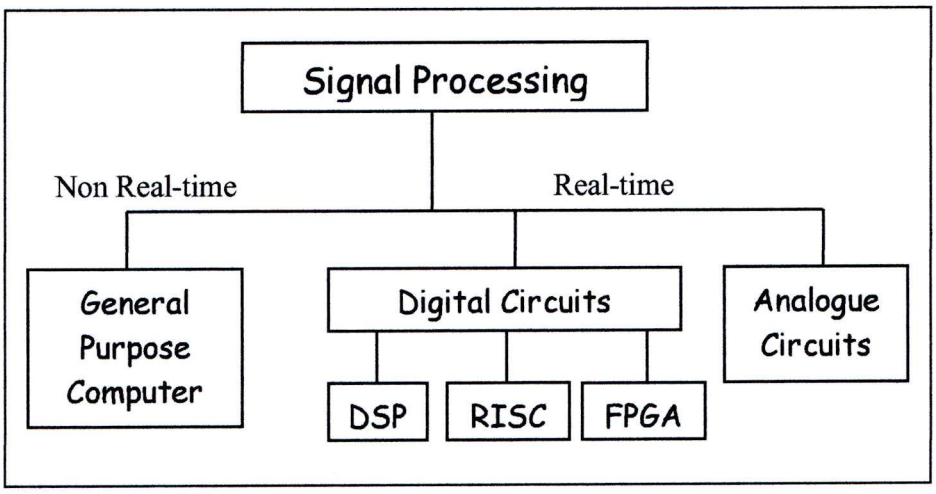

Figure 5.2, Signal Processing Hardware Alternatives.

Referring to figure 5.2, the non-real time alternatives can be discounted since the requirement of a hearing aid is to hear the output from the processed input in real time. This leaves the alternatives of analogue and digital real time circuits. The first generation of hearing aids used analogue technology to implement the required real-time signal processing but developments in high speed DSPs has now digital technology to be used for hearing aid circuits. Digital implementations have been introduced to yield improved advanced signal processing and noise reduction. The other alternative hardware approach is to use a hybrid system using various continually restored analogue data on a minimal number of wires which can be passed and processed like digital data [5.4]. This hybrid approach can use the advantages each of digital and analogue techniques to create the best solution for programmability, power

consumption and area consumption. The development of a hybrid approach is discussed further in section 5.3, but it can be seen that a hybrid approach can provide an efficient solution for the key considerations and specifications for a hearing aid design as highlighted in chapter 4.

5.2.1 Current Hearing Aid design: DSP and Microcontroller

The majority of current digital hearing aids use an architecture that incorporates a traditional DSP processor and a control processor to perform other non-signal processing functions. This has significant benefits in terms of power efficiency and total system cost, but requires parallel memory systems, duplication of processing functionality, and communication between the processors can be inefficient [5.5]. Figure 5.3 shows the die layout for a digital hearing aid implemented with this hardware approach. This ASIC approach incorporates: an analogue pre-amplifier, ADC, DSP, DAC and a co-processor controlled EEPROM for filter parameters, gain control and automatic gain control. The die is approximately 30 mm² in 0.8 µm CMOS technology and contains 30,000 gates with 120,000 transistors.

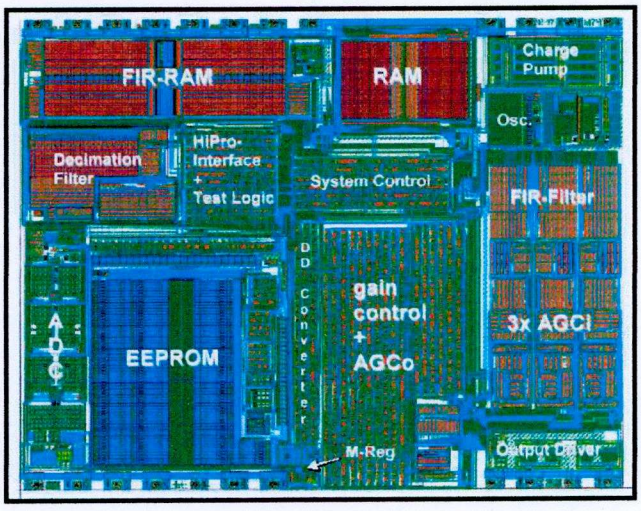

Figure 5.3, TÜRK+TÜRK ELECTRONIC GMBH
Digital Hearing Aid Die Layout (taken from [5.6]).

Referring back to the specification for a hearing aid design presented in chapter 4, this ASIC approach does have some limitations. Firstly the power consumption of an ASIC, although good, cannot be described as

ideally efficient because of the number of transistors required and power consumption used for memory and control of data flow. Secondly, flexibility is limited because of the specifically designed nature of the ASIC and because of the way a significant amount of effort has been used to precisely fit the design into as small as possible circuit package. This precise design has obvious cost implications because the design is very specific to single hearing aid device.

5.2.2 A fresh approach: Digital vs analogue?

The specification in chapter 4 highlights the important need for miniaturisation in a hearing aid circuits and systems. Therefore VLSI technology would be needed to implement the circuits. What is less clear is whether the design should be analogue VLSI or digital VLSI technology.

For many years both digital and analogue hearing instruments have been available. The initial reason for the use of a digital approach is the performance requirements. Signal processing in general, and particularly audio processing, requires high levels of mathematical complexity. The analogue domain is impractical for implementing these applications. However in the digital domain, signal processing technology has been developed to implement these higher mathematical applications in an effective, cost and power efficient way. Digital signal processing also has reliability advantages over analogue processing. Digital technology is significantly more consistent and stable with no errors or drifting of the output with time, variations in temperature, voltage instability or variations in components. However as discussed in the next section, it can be shown that a digital hardware approach for modelling a biological system is not the most efficient method [5.4]. Since our hearing aid design has been based on the natural human ear, this indicates that a purely digital approach may not be the most efficient hardware approach.

5.3 Digital vs. Analogue for VLSI circuits

Research into using electronics to model neurobiology has reviewed the pros and cons of analogue and digital computation for VLSI circuits. This work is obviously highly relevant as a foundation for the hardware approach for a Hearing aid based on the natural human cochlear.

If we start at the most basic level, analogue circuits compute using continuous values of physical variables. These variables can either be current or voltage and are restricted to a given range. It is more typical to use voltage as the variable. The voltage variable range is ultimately limited between the upper limit of the power supply and the lower limit which is determined by the level of noise which is intrinsic to the circuit.

On the other hand digital circuits use discrete values of physical variables to typically denoted 0, for a lower limit voltage, and 1, for an upper limit voltage. The precision of the physical variable used to represent the discrete value is less important in digital so a range of voltages can be used to represent each discrete value, i.e. 0 or 1. This means fundamentally digital circuits are more tolerant of intrinsic noise from transistor physics and thermal fluctuation.

The use of physical parameters for analogue circuits means the circuits are designed using the physical relationships of transistors, capacitors, resistors, using voltage and Kirchoff's current laws. Digital computation uses Boolean maths and logic relation with AND, OR, NOT, NAND and XOR primitives. These primitives are basically constructed form transistors acting as switches. Thus for digital circuits we can see the amount of computation which is performed by a single transistor is low, whilst for an analogue circuit the amount performed by a single transistor is high.

As well as the amount of computation performed by a single transistor the amount of information presented by a single wire at any given time is greater for analogue circuits. Within a digital circuit one wire is needed

for each bit of information at any given time, but for an analogue circuit one wire can represent many bits of information at a given time. This property is fundamental to the idea of analogue as having advantages in terms of efficiency compared to digital. For example the addition of two parallel 8-bit numbers takes one wire in analogue circuits, (using Kirchoff's current law), whereas it takes 240 transistors using a cascade of 8 full adders in CMOS digital circuits [5.4].

It has been shown that the Lyon Mead model we are using as a frequency feature extraction system of a hearing aid can produce significant power savings when comparing an analogue vs. digital large scale implementation of the silicon cochlea. [5.7]. Depending on how the digital circuit is implemented the advantage in power consumption can range from a factor of 300 to 1×10^5.

There are other considerations when looking at the effective efficiency of analogue vs. digital. For analogue circuits the use of physical parameters means that computation is highly sensitive to noise and prone to offsets from mis-matches in physical parameters. In digital circuits noise results in round-off errors, but the computation is not affected by mis-matches in physical parameters. However a single bit error can cause catastrophic failure whereas in analogue computation errors can be less catastrophic and degradation of performance is more gradual.

Looking at an overall hearing aid system the basic idea is that analogue circuits are fundamentally more efficient because of their use of physical parameters, but they are highly limited by noise. However digital circuits divide the information into bits and each bit is represented on a single wire and therefore the effect of noise on a each wire has a diminished effect on the overall computation. Also complex analogue circuits tend to accumulate noise because of the lack of signal restoration along the circuit, whereas complex digital circuits tend to incorporate signal restoration at each stage. It has therefore been suggested that the most effective and efficient approach would be a hybrid approach combining the advantages of continuous signal restoration with continuous-time analogue

computation. [5.4]. It is believed that this hybrid approach is, fundamentally, how highly efficient biological systems operate [5.8]. Since we are using the biological ear as a template for a hearing aid it seems obvious to use the biological system as an inspiration. For example the cochlea is essential to an analogue system that carries out computation on one wire but encodes and distributes information that can then be continually restored and used by a digital/nervous system on many wires.

5.4 Conclusion

This chapter studies the way in which the circuits and systems of a hearing aid can be implemented in hardware using the specification detailed in chapter 4. As previously discussed, this thesis concentrates on implementing the frequency feature extraction stage of a hearing aid which is inspired by the cochlea structure within the natural human ear.

The chapter highlights the Lyon & Mead model as an electronic model of the human cochlear which in affect achieves the frequency feature extraction stage. The idea is to use this model to implement a cascade of filters which can be used to tap off the frequency components at each stage achieving the frequency feature extraction stage within a hearing aid. A typical feature extraction stage would have 15 channels with outputs arranged linearly on a log scale ranging from 10 Hz to 18 kHz.

The complete range of different hardware approaches for the signal processing are initially highlighted and discussed in terms of their advantages, disadvantages and how they have been previously used for current hearing aid systems. The conclusion of this is that the best approach to fit the specification given in section 4.3 is to implement the system as a VLSI ASIC. It is less easy to conclude whether a digital or analogue approach is better for the systems and circuits. Therefore the chapter discusses the fundamental difference between a digital versus analogue approach for VLSI. It is further concluded that a hybrid digital and analogue approach has been shown to be most efficient for these types

of circuits. This hybrid approach is also believed to closely match the natural, and extremely efficient, biological approach we see within the natural ear. Therefore as previously discussed, as the hearing aid has to fit within and compensate for deficiencies within the natural ear, it would appear sensible to take inspiration from the efficient natural biological hardware approach.

To conclude this chapter a hybrid analogue/digital hardware approach will be used for the hearing aid circuits and systems. For the frequency feature extraction stage this would involve a cascade of analogue filters with channels at ranging frequencies from 10 Hz to 18 kHz. The analogue frequency feature extraction circuit would carry out computation on one wire but distribute the signal processed information, that can then be continually restored on many wires, and subsequently be used by a digital stage.

References

[5.1] Lyon, R. F., C. Mead, C., "An analog electronic cochlea", IEEE Trans. Acoust., Speech, Signal Processing, vol. 36, pp. 1119-1134, Jul. 1988.

[5.2] Kennedy, H. J., Evans, M. G., Crawford, A. C., Fettiplace1, R., "Depolarization of Cochlear Outer Hair Cells Evokes Active Hair Bundle Motion by Two Mechanisms", The Journal of Neuroscience, vol. 26, no. 10, pp. 2757-2766, Mar. 2006.

[5.3] Lyon, R. F., "Analog implementations of auditory models", Proc. of the workshop on Speech and Natural Language, pp. 212-216, Feb. 1991.

[5.4] Sarpeshkar, R., "Analog versus digital: extrapolating from electronics to neurobiology", Neural Computation, vol. 1, pp. 1601-1638, 1998.

[5.5] Schneider, T., Brennan, R., Balsiger, P., Heubi, A., "An Ultra Low-Power Programmable DSP System for Hearing Aids and Other

Audio Applications", Proc. of International Conference on Signal Processing Applications and Technology (ICSPAT '99), Nov. 1999.

[5.6] Turk + Turk Electronic GMBH, Am Dannekamp 15, D-51469 Bergisch Gladbach, Germany.

[5.7] Sarpeshkar, R., Lyon, R. F., Mead, C. A, "A Low-Power Wide-Dynamic-Range Analog VLSI Cochlea", Analog Integrated Circuits and Signal Processing, vol. 13, pp. 123-151, 1997.

[5.8] Mead, C., "Neuromorphic electronic systems", Proceedings of the IEEE, vol. 78 , no. 10, pp. 1629–1636, Oct. 1990

Chapter 6

6 SOI CMOS FOR ANALOGUE HEARING AID CIRCUITS

Chapter 5 concluded that the most efficient hardware approach for the frequency feature extraction stage of a hearing aid would be to implement a cascade of analogue filters based on the Lyon & Mead cochlea model [6.1]. This would be part of a hybrid VLSI digital/analogue approach for the overall hearing aid system. The hybrid approach [6.2] would use an analogue frequency feature extraction circuit that would carry out computation on one wire but distribute the signal processed information, which can then be continually restored, on many wires, and subsequently be used by a digital stage. Therefore this chapter discusses VLSI CMOS for low power analogue circuits and discusses silicon-on-insulator technology as an appropriate technology for a hybrid CMOS design.

Chapter 6 introduces the CMOS hardware implementation of low power VLSI analogue circuits with particular reference to analogue filters for hearing aids. The case for subthreshold/weak inversion operation is made and discussed. The chapter continues with the introduction of the use of SOI technology for analogue circuits and discusses how SOI devices can improve performance compared to Si bulk devices. The other reason for choosing SOI technology for hearing aid circuits was the ability to use the technology for mixed signal or hybrid circuits/systems and as a platform for microelectromechanical systems (MEMs) such as the microphone to create an almost truly complete system-on-chip (SoC).

This chapter also discusses models that can be used to design and simulate SOI circuits for this hearing aid circuit application. The modelling of SOI also allows a numerical comparison of SOI versus Si bulk devices.

6.1 Introduction

The hearing aid circuit design discussed in this thesis concentrates on the frequency feature extraction system. The circuit was designed using the specification developed and discussed in 4.3. Particular concern was placed on the specification concerning low power consumption, flexibility and size.

6.1.1 Frequency Feature Extraction System

As previously discussed in chapter 5 the Lyon & Mead model [6.1] will be used as the basis to implement a frequency feature extraction system that has been highlighted in the proposed hearing aid design shown in figure 2.4. The frequency feature extraction system will therefore use a cascade of analogue filters to process the input signal and extract data into a number of analogue data wires. Each analogue wire will represent a different frequency tap. The use of a number of wires that can be individually processed has similarities to a digital process; hence the system is described as a hybrid system [6.2]. Therefore the major hearing aid circuits required for the frequency feature extraction system are analogue filters with variable cut–off frequencies that can process the input signal into the frequency taps. Hence we need to look at the development of low power analogue circuit technology.

6.1.2 CMOS VLSI technology

The need for more portable electronic devices has led to the move towards very large scale integration (VLSI). Now total system-on-chip (SoC) designs has led to the need for circuits with both analogue and digital circuitry to exist on the same technology. These, so-called mixed signal or hybrid system devices, integrate circuits formerly partitioned along analogue digital boundaries [6.3]. CMOS technology has been the technology of choice for mixed signal implementations because of its high density, flexibility on the digital side and power saving on the analogue side [6.4]. In recent years research focused on CMOS technology has become increasingly important for analogue systems. Technology has also

moved towards very low power electronic circuits and systems, primarily to limit battery size, weight and time between replacement or recharging of portable electronic devices. In recent years this has led to an increase in research and development activities into low power and low voltage analogue CMOS circuits. Additionally electronic devices produced in submicron CMOS processes cannot be operated with supply voltage above 3 V, forcing the design of corresponding electronic circuits to be low voltage and hence low power.

6.2 Low Power Analogue CMOS

Power consumption in digital CMOS circuits was considered first since it is the commonly known case. The most efficient way to reduce the power consumption of digital circuits is to reduce the supply voltage, since the average power consumption of CMOS digital circuits is proportional to the square of the supply voltage. Discounting leakage currents within devices the power consumption or the heat dissipated, P, is given by,

$$P = C_L V_{dd}^{\ 2} f \tag{6.1}$$

where C_L represents the load and parasitic junction capacitances, f the switching/clock frequency and V_{dd} is the power supply voltage.

The rules for analogue circuits are quite different from those applied to digital circuits. Decreasing the supply voltage unfortunately does not reduce the power consumption of analogue circuits mainly due to the fact that the power consumption of analogue circuits at a given temperature is basically set by the required signal-to-noise ratio (SNR) and the frequency of operation (or the required bandwidth). In addition to these fundamental limits there are also some practical limits and additional obstacles to power reduction.

6.2.1 Fundamental limit of Power in Analogue CMOS

Power is consumed in an analogue circuit to maintain signal energy above the energy of intrinsic noise within the circuit. Intrinsic noise in analogue

CMOS is discussed in depth later in section 6.5. In order to make the signal easily discernable from the noise the energy in the signal must be set to provide a minimum signal-to-noise ratio (SNR) and the required circuit bandwidth.

In order to consider the analogue power consumption of different CMOS technology a simple typical analogue circuit that realizes a single pole at a bandwidth frequency can be considered. From this the power consumed to realize a single pole can be used as a figure of merit [6.4]. This can be achieved by considering the basic integrator circuit shown in figure 6.1, assuming the transconductor is 100% efficient so that all the current pulled from the power supply appears at i_{out} and is used to charge C.

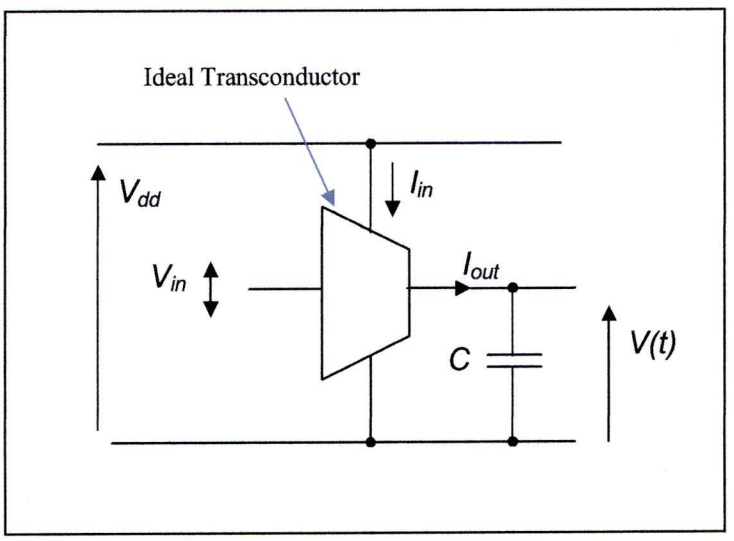

Figure 6.1, Basic integrator circuit, assuming the transconductor is 100% efficient.

The power (V_{dd}xI_{in}) to create the pole consumed from the power supply, V_{dd}, which is required to give a sinusoidal voltage $V_{out}(t)$, at a frequency f, across capacitor C with a peak-to-peak signal voltage amplitude V_{in} is given by [6.5],

$$P = 8kT.f.SNR.\frac{V_{dd}}{V_{in}} \tag{6.2}$$

where k is the Boltzman constant and T is the temperature.

From (6.2) it can be seen that the power consumption of an analogue circuit at a given temperature is set by the required minimum SNR and the frequency of operation, i.e. the bandwidth required. Also to achieve a power efficient analogue circuit it can be seen from (6.2) the ratio of supply voltage and signal peak-to-peak amplitude that the voltage swing of the signal should be maximised so that $V_{in} = V_{dd}$. Due to overdriving of the circuit it is not possible to have $V_{in} > V_{dd}$.

6.2.2 *Practical limits of Power Consumption in Analogue CMOS*

As well as fundamental limits there are also practical limits that are technology dependant. These can be seen as obstacles or technological limitations to power consumption reduction in practical circuits. There are a number of different ways of overcoming many of these limitations at a device to system level.

(I.) Capacitors and additional parasitic capacitance increase the capacitance necessary to achieve a given bandwidth. Therefore reducing the parasitic capacitance can reduce the analogue circuit power consumption.

(II.) An increase in intrinsic noise from the circuit equates to an increase in power consumption as the SNR in (6.2) is increased. Therefore sources of noise must be minimised.

(III.) Analogue circuit design often requires bias current to be used. The power used in setting up bias currents is wasted and must be minimized. Inefficient biasing schemes can also be noisier and therefore increase power consumption.

(IV.) The need for precision and matching within devices to attain good functionality leads to larger dimensions for components which results in increased parasitic capacitance and power. Therefore the ability of a technology to achieve well matched small scale dimensions is important.

(V.) From (6.2) the power is increased if the signal at any node which corresponds to a pole location within the bandwidth has a peak-to-peak voltage smaller than the supply voltage. Therefore circuit design should amplify the signal as early as possible to its maximum possible voltage and maintain this level as much as possible.

(VI.) The power required for the capacitive load, including parasitic capacitors, is supplied by current I_{out}. Therefore I_{out} needed to obtain a given bandwidth is inversely proportional to the ratio of the device transconductance and drain current, g_m/I_D, for the active devices,

$$I_{out} = \frac{I_D}{g_m} 2\pi C_L f \tag{6.3}$$

where I_D is the drain current, g_m is the transconductance and f is the frequency. Therefore minimum power consumption is achieved when g_m/I_D is maximised. As discussed later this is achieved in MOSFET devices when they operate in weak inversion, section 6.4, and using SOI technology devices, section 6.6.

6.2.3 Other Obstacles to Low power analogue CMOS design

In addition to the fundamental and practical limitation for power consumption reduction there are also some other historical and modelling obstacles.

(I.) Often analogue circuit designers use well known and well used analogue building blocks from existing libraries and these are not necessarily power efficient or compatible with low power designing.

(II.) The use of very low bias currents is often avoided because of the lack of transistor models that characterise correctly the devices at low current levels. However models such as the EKV model, which is described further in section 6.7.8, operate in all regions of transistor operation from weak inversion to strong inversion and therefore can model low current levels.

6.2.4 Supply voltage reduction

In order to keep transistors operating in saturation the drain–source voltage must be above a given saturation voltage V_{sat}. Therefore, there is a limit on how low the voltage drop across each device can be and within a circuit a number of transistors can be stacked in series between the supply rails and this, therefore, limits how low the supply voltage can be. V_{sat} is significantly lower in weak inversion than strong inversion as discussed in section 6.3, 6.4.

The previous discussion shows that beyond the fundamental limits there are many problems that can arise in low voltage/low power analogue CMOS circuit design. These problems are often directly related to the properties of MOS transistors at a device level and in order to overcome and understand them it is important to design circuits using correctly modelled transistors even at very low currents.

It can also be noted, the trend towards mixed signal or hybrid implementations has made it necessary to develop MOSFET device models that can be easily adapted to analogue as well as digital circuits [6.6, 6.7].

The first stage to designing an analogue CMOS circuit is to understand the device level operation of CMOS at the large signal scale in the different regions of inversion, i.e. strong and weak inversion. This then has important implications in terms of circuit design and power consumption.

6.3 The MOSFET Transistor in strong inversion

6.3.1 Large Signal Analysis

In strong inversion the inversion layer charge, Q_I, is much greater than the depletion layer doped charge, Q_B, i.e. $|Q_I| >> |Q_B|$.

This means in strong inversion Q_I is proportional to the effective voltage $(V_{GS}-V_{TO})$, where V_{GS} is the gate–source voltage and V_{TO} is the threshold voltage,

$$Q_I \cong -C_{ox}(V_{GS} - V_{TO}) \tag{6.4}$$

where C_{ox} is the gate oxide capacitance, which is dependent on the oxide thickness and dimensions of the device.

The transistor is saturated in strong inversion if $(V_{GS}-V_{TO})<V_{DS}$ and in saturation the drain current is given by,

$$I_D = \frac{1}{2}\mu_c C_{ox}^{'} \frac{W}{L}(V_{GS} - V_{T0})^2 \tag{6.5}$$

where μ_c is the surface mobility of the channel and W, L are device dimensions.

6.3.2 *Small Signal Analysis*

The large signal MOSFET equations described in (6.5) are inherently non-linear and hence within a circuit these devices would represent a system of non-linear equations. Systems such as filters and amplifiers are linear circuits and require linear behaviour to describe them. For a MOSFET device this can be achieved by looking at the small signal response of devices. In simple terms we are zooming in on nonlinear curves until they appear linear over a small range. Thus if we bias a transistor around a certain operating point and then only make small changes around this point we can approximate the device as having a linear behaviour. This is called small signal analysis.

The most important small signal parameter of a MOSFET is the transconductance, g_m. The transconductance defines the relationship between the gate to source voltage, V_{GS}, and the drain current I_D.

$$g_m \equiv \frac{\partial I_D}{\partial V_{GS}} \tag{6.6}$$

From (6.5) we can express the transconductance of an above-threshold MOSFET in saturation as

$$g_m = \sqrt{2\mu_c C_{ox} I_D \frac{W}{L}} \tag{6.7}$$

6.3.3 *Channel length modulation*

At a simple level in saturation after pinch-off the device drain current is taken to be constant. At a more complex level this is shown not to be true. This is because the depletion region around the source and drain is affected by the voltage of the source and drain. The effect of increasing the drain voltage in the saturation region widens the depletion region around the drain. This in turn has the effect widening the pinch-off region and thus shrinking the channel by a small amount. The effect is called "channel length modulation". A shrinking channel increases the current through the device. In short-channel devices (length near the minimum length allowed by the technology process), the effect on current is very complex and difficult to model. Most MOSFET models use empirical models to fit observed behaviour with first-order or higher-order models. This issue is complicated by such effects as "drain-induced barrier lowering" (DIBL) where in short devices the drain can act as an additional side-gate coupling to the channel through the drain depletion capacitance. This can lead to a dramatic increase in drain current as drain voltage increases.

A simple first-order model for channel length modulation uses a parameter called the "early voltage", V_A. The early voltage parameter is taken from plots of I_D versus V_{DS} for several values of V_{GS} extrapolating the slope of the saturated current. The curves tend to converge on the V_{DS} axis at the same point, V_A. Since longer channels are affected less by the modulation effect, $V_A \propto L$, (for long-channel devices). Thus the current in the saturation region for a device can be modified to include a first-order term of V_{DS}.

Equation (6.5) now becomes,

$$I_D = \frac{1}{2}\mu C_{ox}^{'}\frac{W}{L}(V_{GS} - V_{TO})^2\left(1+\frac{V_{DS}}{V_A}\right) \tag{6.8}$$

From (6.8) we can define a small-signal parameter called output resistance to describe the effect,

$$r_o \equiv \frac{\partial V_{DS}}{\partial I_D} = \frac{V_A}{I_D} \tag{6.9}$$

6.4 The MOS transistor in Weak Inversion

Micropower VLSI circuits operate at a supply voltage of typically 1 to 3 V and very low currents of typically 1 nA to 1 μA [6.8]. Historically the field of micropower circuits began in the late 1960's with the development of electronic watches and has since grown and expanded with the ever increasing need for portable electronic equipment, medical devices and telecommunication devices. Very low power consumption is also needed for devices powered solely by solar or electromagnetics [6.9]. In order to achieve low powered CMOS circuits we need to think of devices operating at these low current levels. This involves operating the devices within the weak inversion region.

The models discussed earlier predict no current flow through a device below the threshold voltage as shown in figure 6.2a. However this is not the case and when the gate-source voltage is below the threshold, i.e. operating in subthreshold, it is well known that the channel current decreases approximately exponentially as shown in figure 6.2b. This occurs when the channel is said to be weakly inverted and this region of operating is called "weak inversion" as opposed to "strong inversion". Between the strong and weak inversion regions there is a third region called "moderate inversion". Within the weak inversion region the current continues to reduce exponentially until there is no inversion in the channel. In fact there is an operating limit at the lowest current levels created by intrinsic random noise current inherent within the channel. The noise is discussed further, later in this chapter.

In weak inversion the inversion layer charge is much less than the depletion region charge, i.e., $Q_I << Q_B$ in weak inversion.

For gate voltages less than the threshold voltage, V_{TO}, (i.e. subthreshold or weak inversion region) the drain current obeys an expediential relationship of the form,

$$I_D \propto \exp\left(\frac{V_{GS}}{nU_T}\right) \tag{6.10}$$

where U_T is the thermal voltage given by,

$$U_T \equiv \frac{kT}{q} \cong 26 \text{ mV at room temperature},$$

n is a parameter greater then 1 and $V_S = 0\text{V}$

The parameter 'n' is called the body effect coefficient. The value of 'n' influences both the drive current and subthreshold swing of a MOSFET [6.10]. Appendix A shows the theory for deriving an expression for 'n' the body effect coefficient from the oxide capacitance C_{ox} and the depletion capacitance C_{dep} between the gate voltage and the bulk voltage, V_G and V_B, of an Si bulk MOSFET device as illustrated in figure 6.3, where the surface potential is shown, ψ_s.

$$n = 1 + \frac{C_{dep}}{C_{ox}} \tag{6.11}$$

The value of 'n' influences both the drive current and subthreshold swing of a MOSFET [6.10]. Since the depletion capacitance is virtually constant in the weak inversion region, 'n' is considered to be constant. The value of 'n' is slightly different for PMOS and NMOS, but n = 1.4 is taken to be a good approximation.

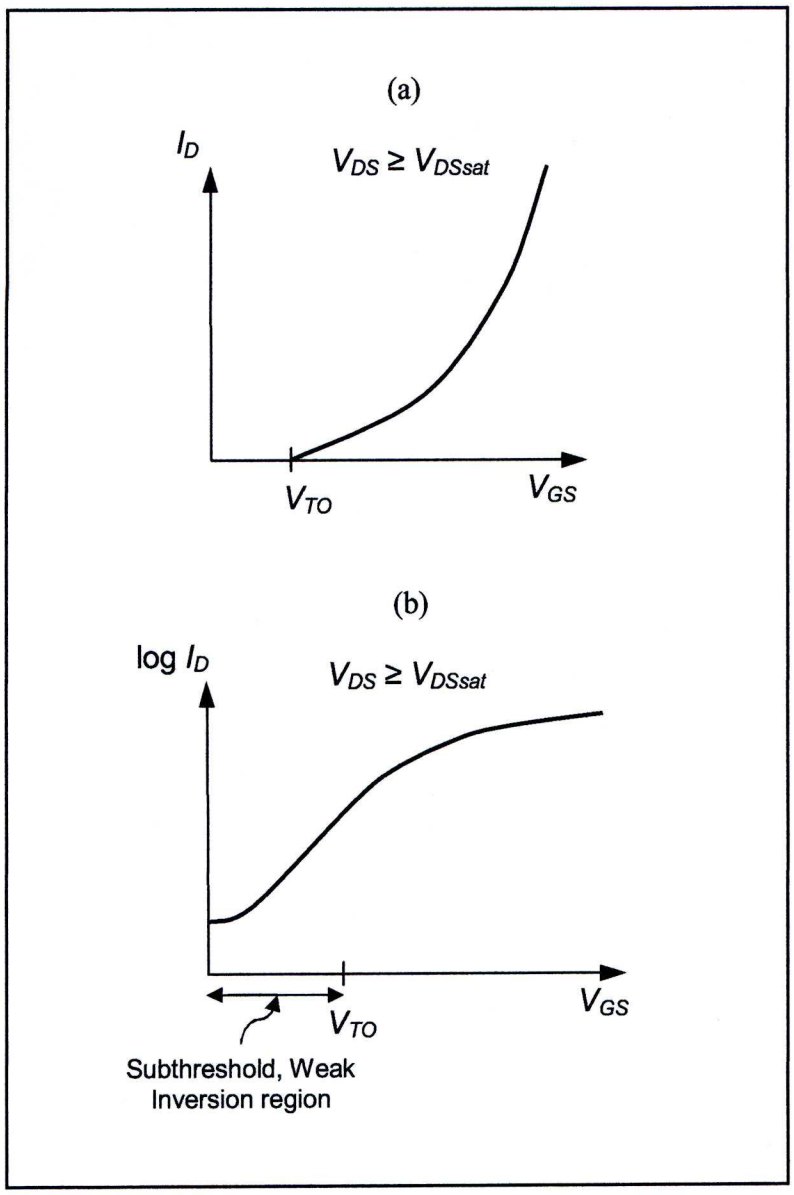

Figure 6.2, a) Strong inversion model for I_D versus V_{GS} for a device operating in saturation ($V_{DS} \geq V_{DSsat}$), showing no drain current below V_{TO}; b) Drain current below V_{TO} for subthreshold/weak inversion for I_D versus V_{GS} for a device operating in saturation ($V_{DS} \geq V_{DSsat}$).

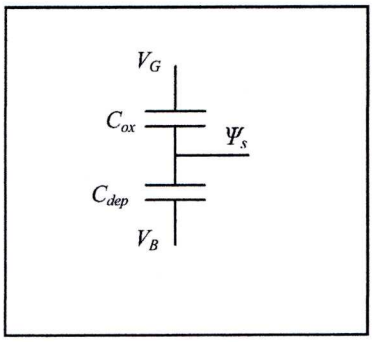

Figure 6.3, The body effect coefficient Si bulk MOSFET.

We have already seen $I_D \propto \exp(V_{GS}/nU_T)$ for $V_s = 0$ V so we can now say the charge concentration (at the channel surface) at the source ($x = 0$) and at the drain ($x = L$) is given by,

$$\begin{aligned} |Q'_{I0}| &\propto \exp\left(\frac{V_S - V_G/n}{U_T}\right) \\ |Q'_{IL}| &\propto \exp\left(\frac{V_D - V_G/n}{U_T}\right) \end{aligned} \tag{6.12}$$

The concentration of electrons decreases linearly from the source to the drain resulting in a constant concentration gradient. Hence from the charge at either end of the channel we can write an expression for the drain current,

$$I_D = -\frac{W}{L}\mu\, U_T (Q'_{I0} - Q'_{IL}) \tag{6.13}$$

Using (6.12) above this leads to,

$$I_D = I_0 \frac{W}{L}\exp\left(\frac{V_G}{nU_T}\right)\left[\exp\left(\frac{-V_s}{U_T}\right) - \exp\left(\frac{-V_D}{U_T}\right)\right] \tag{6.14}$$

where I_0 is a process dependent constant. For NMOS,

$$I_{0n} \equiv 2\mu_n C_{ox} \frac{W}{L} nU_T{}^2 \exp\left(\frac{-V_{T0n}}{nU_T}\right) \tag{6.15}$$

Typically I_{0n} is in the range of approximately 1×10^{-15} A to 1×10^{-12} A, similar to (6.15), for PMOS I_{0p} is approximately 1×10^{-19} A to 1×10^{-14} A. Rearranging (6.14) we can get,

$$I_D = I_0 \frac{W}{L} \exp\left(\frac{V_G - nV_S}{nU_T}\right)\left[1 - \exp\left(\frac{-V_{DS}}{U_T}\right)\right] \tag{6.16}$$

We can note that when $\exp(-V_{DS}/U_T) << 1$, the last term is approximately equal to one and therefore this exponential term can be ignored. Within a 2% error $\exp(-V_{DS}/U_T) << 1$ occurs for $V_{DS} > 4U_T$, since $e^{-4} = 0.018$. This provides the criteria for saturation in weak inversion and (6.16) simplifies to,

$$I_D = I_0 \frac{W}{L} \exp\left(\frac{V_G - nV_S}{nU_T}\right) \text{ for } V_{DS} > 4U_T, \text{ (saturation)} \tag{6.17}$$

At room temperature, $4U_T \sim 100$ mV, so in order to keep a weak inversion region device in saturation $V_{DS} > 100$ mV is required. This fact is very advantageous for a low voltage, low power circuit because of the low value of V_{DS} needed to keep a device in saturation.

6.4.1 Transconductance in weak inversion

The transconductance of a weak inversion device is easy to derive from (6.17).

$$g_m = \frac{I_D}{nU_T} \text{ (in weak inversion)} \tag{6.18}$$

It can be noted this is independent of device geometry, (W and L).

6.5 Noise in analogue circuits

Having modelled the operation of a CMOS transistor in analogue circuits we now need to consider noise in analogue circuits. Intrinsic noise is a statistic measure and not an exact continuous time voltage or current. In order to gain useful analysis noise is viewed as a continuous time source. For practical analysis this enables worse case noise to be considered.

However it is always important to remember the random nature of noise. This is because if we are trying to alter the effect of noise by say, doubling the size of a transistor, it does not directly follow that the noise will be directly doubled, because random noise in one half of the transistor is not going to randomly occur at exactly the same time as random noise in the second half of the transistor.

The noise phenomena considered here are caused by small current and voltage fluctuations that are generated within the devices themselves. This chapter does not consider extraneous pick-up of noise from external sources that can also be a problem in analogue hearing aid circuits. The existence of noise is due to the fact that the electrical charge is not continuous but is carried in discrete amounts equal to the charge of an electron and is associated with the fundamental processes in semiconductor components. Hence the steady current I externally observed, in fact, comprises of a large number of random independent current pulses. In essence, noise is a random variable that is treated as having a density function.

The study of noise represents a **lower** limit for operating electrical signals below which these signals cannot be amplified without significant deterioration in quality. This is because the operating signal cannot be distinguished from the noise. In a simple transistor noise can be modelled by a current source in parallel with the I_D current. The current source can be translated to an equivalent noise voltage input using the transistor's g_m. This current source models several sources of noise that are important in CMOS components.

6.5.1 Shot Noise [6.11]

Shot noise is associated with a direct-current flow across a PN junction. The passage of carriers across the junction can be modelled as a random event and is dependent on the carrier having sufficient energy and direction towards the junction. The random fluctuations in I_D are termed

shot noise and are given in terms of a mean-squared variation about the average value. Thus the shot noise current is given by,

$$\overline{i_{sn}^{\,2}} = 2qI_D\Delta f \tag{6.19}$$

where q is the electronic charge, I_D is the average current and Δf is the bandwidth.

6.5.2 *Thermal Noise[6.12]*

Thermal noise is generated by the random thermal motion of electrons and is unaffected by the presence or absence of direct current. As the motion of electrons is related to the absolute temperature T, thermal noise is directly proportional to T, and as T approaches zero, thermal noise approaches zero. The potential and therefore current generated by thermal noise is given by,

$$\overline{e_{th}^{\,2}} = 4kTR_{th}\Delta f \tag{6.20}$$

therefore,

$$\overline{i_{th}^{\,2}} = 4kT\frac{1}{R_{th}}\Delta f \tag{6.21}$$

where k is the Boltzmann's constant and R_{th} is the resistor or equivalent resistor in which the thermal noise occurs.

In MOSFETs the most important resistance to be considered is the channel resistance. The effective channel resistance can be expressed as $1/(\Gamma_{rch}\times g_m)$, where Γ_{rch} is 2/3 for MOSFETs in the saturation region [6.13]. Therefore (6.21) becomes,

$$\overline{i_{th}^{\,2}} = \frac{8}{3}kTg_m\Delta f \tag{6.22}$$

6.5.3 *Flicker Noise*

Flicker noise or ($1/f$) noise has been extensively studied, as it dominates the low-frequency noise; at the same time there is an increasing need to accurately design low-noise analogue circuits in CMOS technology. Noise like properties with a $1/f$ power law have been observed in practically all electronic materials and similarly in mechanical, biological, geological and

even in musical systems. However there has been no entirely satisfactory explanation for the origins of flicker noise despite extensive study [6.14]. Recent studies point toward an explanation of the flicker noise based on the carrier number fluctuation theory [6.15]; the same theory can also be applied to the FD SOI MOSFETs [6.16]. However there is a competing model based on mobility fluctuation theory, (see later), and the physics behind flicker noise in MOSFETs is still a topic of discussion. It is widely accepted that low frequency noise in MOSFETs is, in both theories, a result of trapping and detrapping of carriers in energy states near the surface of semiconductors [6.17].

Carrier Number Fluctuation theory

Number fluctuation theory was first developed in 1937 for germanium filaments [6.18] and proposed flicker noise was a surface effect. The theory was first applied to MOSFETs in 1968 [6.19, 6.20] and states that noise is caused by the tunnelling of carriers from the channel into traps inside the oxide layer. The MOSFET theory has further been developed to take into account capacitive components for small-signal analysis and to account for transistor operation in all regions from strong to weak inversion [6.21]. Finally the theory has been developed into a compact model [6.22],

$$\overline{I_{cnf}^{\,2}} = \frac{q^2 N_{ot} \mu I_D}{nC'_{ox} L^2} \Delta f \cdot \frac{1}{f} \int_{Q'_{IS}}^{Q'_{ID}} \frac{1}{nC'_{ox} U_T - Q'_I} dQ'_I \tag{6.23}$$

where N_{ot} is the effective number of traps [6.22] and is a technology parameter, μ is the effective mobility, and Q_I is the inversion charge density.

Equation (6.23) is valid for all regions of operation. In weak inversion Q_{ID}, $Q_{IS} << nC_{ox}U_T$ and using a first order expansion it is possible to simplify (6.23) to [6.22],

$$\frac{\overline{I_{cnf}^{\,2}}}{I_D^{\,2}} = \frac{N_{ot}}{WL\left(nC'_{ox} U_T / q\right)^2} \frac{\Delta f}{f} \tag{6.24}$$

In strong inversion $Q_{ID} \approx Q_{IS} \approx C_{ox}U_T(V_G - V_{T0})$ and using a first order expansion it is possible to simplify (6.24) to [6.22],

$$\frac{\overline{I_{cnf}^{\ 2}}}{I_D^{\ 2}} = \frac{qN_{ot}}{WLC'_{ox}(V_G - V_T)^2}\frac{\Delta f}{f} \tag{6.25}$$

From (6.24) and (6.25) it can be seen that in weak inversion the theory predicts that the weak inversion flicker noise will increase with I_D^2 while in strong inversion the flicker noise only increases by I_D.

Mobility Fluctuation theory

Mobility fluctuation theory stated that fluctuations in conductivity are due to fluctuations in mobility and not in the number of carriers [6.23]. This leads to an empirical parameter, "Hooge's constant" taken from experimental data. The theory was developed for MOSFET devices [6.23] using lattice scattering theory [6.24] to give a strong inversion noise of,

$$I_{mf}^{\ 2} = \frac{\mu_0}{\mu_{eff}}\left(\frac{\mu_0}{\mu_l}\right)^2 \frac{q\alpha_H I_D^2}{WLC_{ox}(V_{gs} - V_T)f} \tag{6.26}$$

where μ_o is the mobility, μ_{eff} is the effective mobility, μ_l is the mobility if only lattice scattering were present and α_H is the Hooge's constant.

Combined Models: In an effort to improve SPICE modelling and simulation of circuits the unified model has been developed [6.25, 6.26] to quantify the contribution of surface mobility fluctuations by examining the channel current modulation due to the trapping of a single electron and using a scatter coefficient. This approach gives,

$$\overline{I_{1/f}^2} = \frac{kTI_D^2}{\gamma fWL}\left(\frac{1}{N} + \alpha'\mu\right)^2 N_T(E_{fn}) \tag{6.27}$$

where γ is the attenuation coefficient of electron waves into the oxide; α' is the scattering coefficient and $N_T(E_{fn})$ is the number of traps at the quasi-Fermi level.

Simple Empirical model: Many versions of SPICE use empirical models that are seen as over simplification compared to theoretical and combined models. A common empirical model is,

$$\overline{I_{1/f}^2} = \frac{K_F g_m^{\ 2}}{C_{ox}' WL} \cdot \frac{1}{f} \Delta f \tag{6.28}$$

where K_F is an empirical constant which can be different in weak and strong inversion.

6.6 SOI

We can now introduce silicon–on–insulator technology for MOSFET devices. The main advantages of SOI over conventional bulk technology have been well documented, namely reduced processing stages, lowering of the threshold voltage of the device without increasing off-state leakage current, reduced vertical field at the surface channel thereby increasing device mobility. SOI is making a breakthrough into industrial applications due to the limitations of bulk CMOS.

6.6.1 Overview of the state of SOI technology in electronics

With the advent of systems on a chip, and the increasing need of portable electronic equipment such as laptops and mobile phones, the need for faster and reduced power microprocessors is growing and industry strategy is changing towards SOI. Figure 6.4 compares the basic structure of an SOI transistor and a traditional Si bulk transistor. The fundamental difference is the layer of oxide beneath the device. The thickness of this device film above the isolating oxide divides the SOI transistor into two types; (i) partially depleted (PD SOI), with a deep oxide layer (thick film, normally > 100 nm) or (ii) fully depleted (FD SOI), with a shallow oxide layer (thin film, normally < 50 nm). As the names suggest the difference refers to the amount of the channel above the oxide that becomes depleted as a result of a charge at the gate.

For low voltage low power digital VLSI circuits the advantages of SOI over conventional SI bulk CMOS have been well known and documented [6.27]. The primary advantage of all SOI devices has always been that the dielectric isolation of devices provides reduced parasitic capacitance and leakage currents which enable higher operating frequencies for switching and lower power consumption compared to the junction isolation of conventional Si bulk devices. The lack of a need for wells in SOI circuits to separate N^+ from P^+ regions means that device density can be higher reducing area consumption.

6.6.2 SOI advantages

Compared with Si bulk CMOS SOI has reduced junction capacitance. In bulk MOSFET devices interaction between the device and substrate creates a number of unwanted parasitic capacitances. In SOI technology the parasitic capacitances are significantly reduced because the source/drain diffusion areas couple to the substrate via the oxide layer (ε_r = 4.1) instead of silicon (ε_r = 12).

In addition, Si bulk CMOS structures suffer from latch up problems created by the triggering of a PNPN bipolar thyristor structure, particularly with small device dimensions. In SOI CMOS latchup does not occur because there is no current path between the active device and the substrate so no parasitic bipolar devices exist between MOSFETs. Techniques have been developed to avoid parasitic capacitance and latch up in Si bulk devices. Parasitic capacitance can be reduced by creating local interconnections and placing contacts over the field area, and latchup can be avoided by keeping the N-well as far a away from the source and drain as possible and by using epitaxial substrates. These techniques, however, have significant impact on manufacturing costs and chip area consumption.

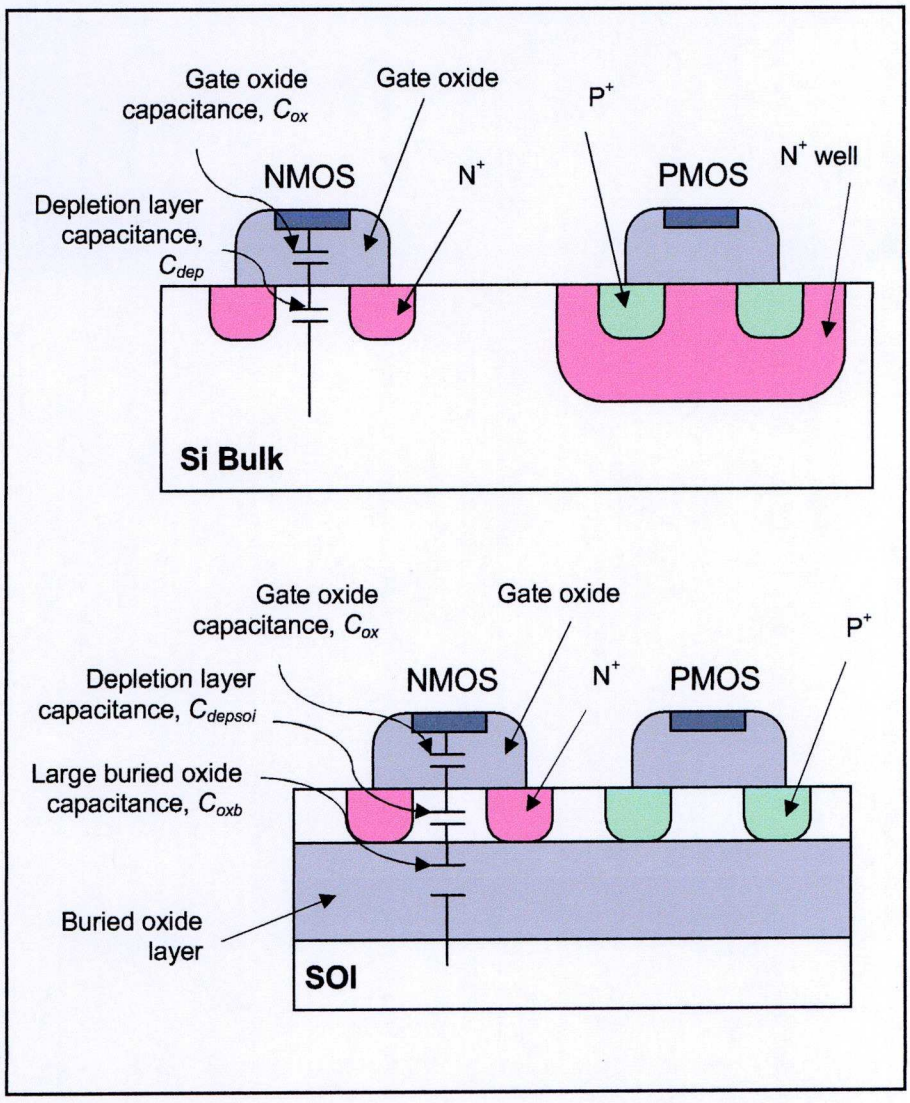

Figure 6.4, Comparison of Si Bulk and SOI device structures.

With the increased reduction in device dimensions over recent years the effects caused by the reduction in channel length in MOSFETs has become an important issue. The reduction of channel length has many effects, as discussed earlier. The so-called "short channel effect" results in a roll-off of the threshold voltage in short channel devices. It is due to the depletion around the source and drain encroaching on the gate controlled depletion zone, resulting in a loss of gate control. The advantage of thin-film, FD SOI is that the channel depth is restricted, therefore the encroachment of the source and drain depletion is limited reducing this problem [6.28]. With less short channel effect it then follows that it is possible to use a lower threshold voltage to improve performance, particularly for low

power, low voltage circuits. This is because the amount of depletion in the channel caused by the gate is directly related to the threshold voltage. Therefore if it is possible to operate a short channel with less gate depletion and the threshold voltage can be lowered.

The body effect coefficient represents the coupling of the gate to the surface potential as shown in figure 6.5 for FD SOI and Si bulk MOSFETs; V_G and V_B are gate voltage and bulk voltages, C_{ox} is the gate oxide capacitance, C_{dep} is the depletion layer capacitance for Si bulk MOSFET, C_{depsoi} is the depletion layer capacitance of the FD SOI layer, C_{oxb} is the capacitance of the buried oxide layer and ψ_s is the surface potential.

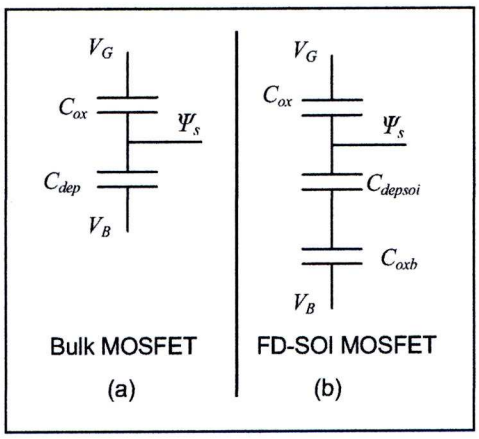

Figure 6.5, The body effect coefficient for, (a), Si bulk and (b) FD SOI CMOS.

Using equations for a capacitive divider and assuming that $V_B = 0$ V, we get,

$$\psi_s = V_G / n \tag{6.29}$$

where the body effect coefficient, 'n', for Si bulk MOSFETS is given by,

$$n = 1 + \frac{C_{dep}}{C_{ox}} \tag{6.30}$$

and for FD SOI MOSFETS is given by,

$$n = 1 + \frac{C_{dep} C_{oxb}}{C_{ox}(C_{dep} + C_{oxb})} \tag{6.31}$$

Thus in FD SOI the body effect coefficient is close to the optimum value of unity, $n \sim 1.05$ to 1.1, compared to $n \sim 1.4$ to 1.6 for Si bulk transistors as discussed in 6.4 [6.29]. The value of 'n' influences both the drive current and subthreshold slope of a MOSFET [6.29].

In MOSFETS another important characteristic which determines the threshold voltage and hence the low power, low voltage limits of a circuit is the "subthreshold slope" of a device. The subthreshold slope determines the minimum voltage swing needed to affect the drain current of the transistor. In digital circuits this is effectively the voltage swing required to turn the transistor on. The characteristic is usually quantified by measuring how many millivolts it takes to change the drain current by one order of magnitude, i.e. one decade of current on a logarithmic scale, and is given in units of mV/decade. The subthreshold swing is limited by thermal voltage, kT/q, which is 60 mV/decade at room temperature. FD SOI thin-film MOSFETS have a sharper subthreshold slope than bulk, again allowing better low power, low voltage performance. The sharper subthreshold slope allows lower subthreshold leakage current for the same threshold voltage compared to bulk CMOS.

FD SOI MOSFET circuits have improved soft error durability from extrinsic radiation due to the thin operating layer.

The complete electrical isolation of circuits with the oxide layer, means the problem of cross-talk noise between circuits on the same chip is reduced. This makes FD SOI an ideal technology for mixed signal SoC design.

The reduced leakage current compared to Si bulk devices is again caused by the sharper subthreshold slope. Thanks to the reduced leakage current SOI has advantages for high temperature applications [6.30]

SOI material has been used to enhance microelectromechanical systems (MEMS) performance for various applications including a piezoresistive silicon pressure sensor [6.31], and capacitive acceleration sensors [6.32]. The development of microphone MEMs would be of particular relevance for hearing aids and with the addition of mixed signal processing, SOI technology presents a viable technology for a complete system–on–chip (SoC) hearing aid. At the time of writing this thesis, to the author's knowledge, there is no development of SOI MEM microphones for hearing aids. However there is published work on CMOS MEM microphones for portable applications [6.33-6.35].

6.6.3 SOI Disadvantages

Figure 6.6 illustrates there is a "kink effect" which appears as a kink in the output characteristic of SOI MOSFETS. The effect is caused by the floating body, which can appear between the channel and the buried oxide layer BOX particularly in PD SOI devices because the channel does not fill all the film of the device. Therefore when the drain voltage is high enough, the channel electrons have sufficient energy to create holes near the drain due to impact ionization. The electrons rapidly move into the channel and drain, while the holes can migrate towards the area beneath the channel creating the floating body potential which is equivalent to a forward bias source–body diode. The increase in body potential can be seen as a similar effect as a substrate bias for a Si bulk MOSFET, in that the change in body coefficient decreases the threshold voltage. This in turn induces an increase of the drain current as a function of drain voltage and hence the kink is observed at a particular level of drain voltage. It is possible to overcome this problem by having a body contact to bias the body of a PD SOI device, however this adds to the cost and complexity of manufacture and increases area consumption.

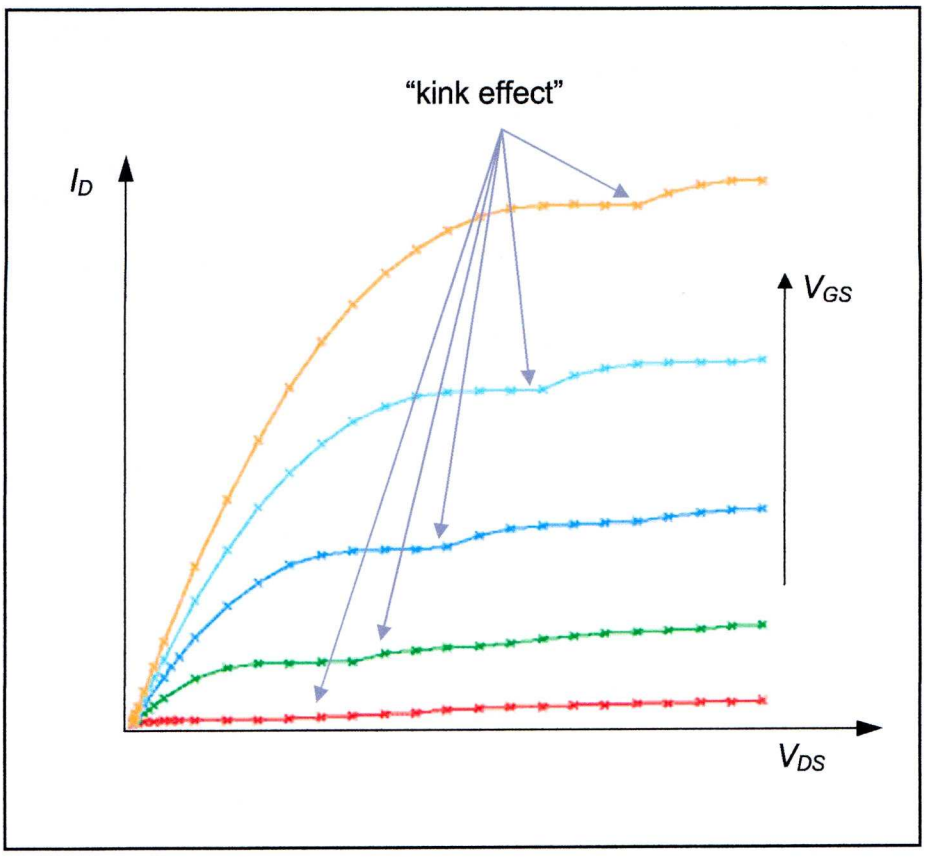

Figure 6.6, Drain current I_D versus V_{DS} for a SOI MOSFET device, showing the "kink effect".

The floating body effect and its apparent reduction of the threshold voltage has a corresponding negative effect on the subthreshold slope, leakage current and noise.

Since the film in FD SOI is much thinner, and as the name suggests, the film is fully depleted across its thickness there is no floating body. Hence despite the difficulties and expense of manufacturing thin-film FD SOI devices, they have obvious advantages in providing a quasi-ideal MOSFET.

However it is important that for thin-film devices full depletion is guaranteed even at low voltages and high temperatures, otherwise the floating body effect occurs in weak inversion operation and the subthreshold slope, leakage current and noise are all degraded.

SOI transistors are thermally insulated from the substrate by the BOX. As a result the removal of excess heat generated within the device is less efficient than in Si bulk devices. This results in an elevation in device temperature. The effect of self heating is a reduction in mobility which affects the device conductivity and subsequent device characteristics. Due to the relatively low thermal conductivity of the buried oxide once sufficient power is dissipated in the device there can be heating up by 50 to 150 ^{0}C and subsequent mobility reduction is observed. It has been shown that the time constant associated with heat dissipated and self heating of SOI transistors is of the order of several tens of nanoseconds. For analogue circuits self heating effects mean the output conductance of the transistor is frequency dependent. At low frequencies self heating effects follow the signal and thus a corresponding reduction in conductance is observed. Self heating in analogue SOI devices also has the effect of increasing transconductance at frequencies above approximately 100 kHz.

As well as self heating, thermal coupling effects can be observed propagating from one device to another, particularly in structures such as current mirrors.

Due to the confined thickness of the film in the SOI layer, devices are less tolerant to electrostatic discharge.

6.6.4 Manufacturing SOI

Commercially, SOI was first used in the 1960s for early satellite and space exploration systems because of its traditional resistance to ionization radiation and the robust voltage isolation of circuits. This was a highly specialised and specific application and hence the high cost of SOI wafer manufacture reflected this. In recent years, work has focused on making SOI fabrication more cost effective so that it can compete with traditional Si bulk technology whilst providing performance benefits. The key material issues are continuity and thickness uniformity of the buried oxide layer (BOX) and thickness uniformity and defectivity of the signal crystal silicon layer above the BOX. Also the interface charge trapped at the

interface of the silicon and BOX must be kept below ~ 1×10^{11} cm^{-2}. The rapidly growing commercial market, particularly in digital circuits, for SOI wafers is continually driving suppliers to improve quality and reduce costs; this will allow the further use of SOI technology for analogue and other circuits.

The most common technique used to manufacture SOI wafers is SIMOX. SIMOX stands for "Separation by Implanted Oxygen". The technique simply involves the formation of a BOX of SiO_2 by implantation of oxygen ions beneath the surface of a standard silicon wafer. This, therefore, makes the technology compatible with many standard Si bulk CMOS processes, which in turn has significant manufacturing cost advantages. However it is still difficult to produce thin-film wafers with low defect density. For this thesis, the circuits were fabricated using the UCL 2 μm FD SOI CMOS technology with 'SmartCut' UNIBOND wafers (compatible with bulk Si CMOS processes), having a buried oxide layer of 400 nm, a Si substrate of 80 nm silicon, 31 nm for gate oxide and a V_{TO} = 0.4 V [6.36]. UNIBOND wafers use two wafers, one of which has hydrogen ions implanted to and thermally grown $Si0_2$ creates the dielectric layer. This first wafer is then bonded at room temperature to another wafer. The wafers are then heat treated in a first phase to 400 to 600 ^{0}C to form a thin monocrystal layer in the second wafer above the BOX. A second heat treatment, > 1000 ^{0}C, strengthens the chemical bonds. Finally the wafer is cut and chemo-mechanically polished to give the necessary surface.

6.6.5 Modelling SOI

There are a number of different modelling approaches for the simulation of thin film FD SOI CMOS circuits. However, as discussed, below for low voltage/low power analogue designs it is essential that we have a continuous model valid from weak through moderate and to strong inversion. This makes many models inadequate and unreliable for analogue designs.

6.6.6 g_m/I_D design process

When considering modelling SOI devices we need to start by thinking about what are the important performance parameters. The g_m/I_D ratio is a very important performance parameter for the design of analogue circuits. Fundamental elements of an analogue circuit, i.e; gain and bandwidth, are strongly related to (g_m/I_D). This is shown in the very simple single transistor circuit of figure 6.7. When a g_m and I_D have been derived from the circuit specification, a device aspect ratio (W/L) is unambiguously determined from the g_m/I_D *vs* $I_D/(W/L)$ curve for a chosen transistor type.

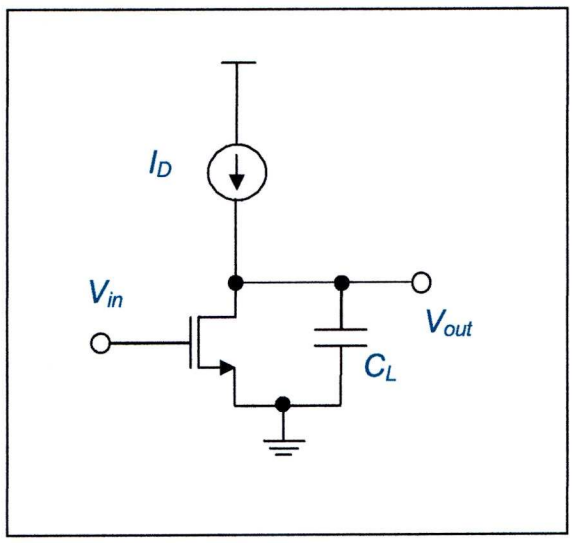

Figure 6.7, Simple CMOS gain circuit.

For the circuit shown in figure 6.7, the open-loop gain (A_{vo}) and the transition frequency (f_t) in terms of g_m/I_D are,

$$A_{vo} = -\frac{g_m}{I_D} V_A \qquad \text{where,} \qquad g_{ds} = \frac{I_D}{V_A}$$
$$f_t = \frac{g_m}{I_D} \frac{I_D}{2\pi C_L} \tag{6.32}$$

6.6.7 g_m/I_D comparison for FD-SOI and Si bulk

The g_m/I_D parameter cannot only be used to design FD SOI circuits but can also be used to compare FD SOI and Si bulk technologies at a design level.

Figure 6.8 shows the comparison of g_m/I_D for similar FD SOI and Si bulk transistors. The increase in g_m/I_D for FD SOI occurs in the weak inversion operating region because of the body effect coefficient. Therefore for micropower designs, the near optimum value of the body coefficient of FD SOI devices, discussed in section 6.6.2 allows g_m/I_D values of 35 V^{-1} compared to only 25 V^{-1} for Si bulk devices [6.37]. This is to be expected from the rearrangement of (6.18) which yields,

$$\frac{g_m}{I_D} = \frac{1}{nU_T} \tag{6.33}$$

The g_m/I_D parameter is fundamental in showing the advantage of FD SOI technology over Si Bulk for an analogue circuit operating with micropower. For example for an OTA operating in weak inversion, the use of FD SOI devices has been shown to simultaneously increase gain and decrease DC power dissipation by ~ 40 %, whilst also the bandwidth can be increased for a given gain and area [6.37]

6.6.8 EKV

The normalized transconductance-to-current ratio is a coherent characteristic that can be used in static, dynamic and noise models. To utilise g_m/I_D in modelling it is important to have an analytical model for devices that provides a continuous representation in all regions of inversion from strong inversion down through moderate inversion and right down to weak inversion device operation. This would be very important when designing and exploiting the power consumption advantages of operating down in the weak inversion region. Examples of such continuous models are EKV [6.7] and the ACM model. For this research the EKV model was used. Figure 6.9 shows the drain current characteristics for all regions of an NMOS modelled by the EKV model.

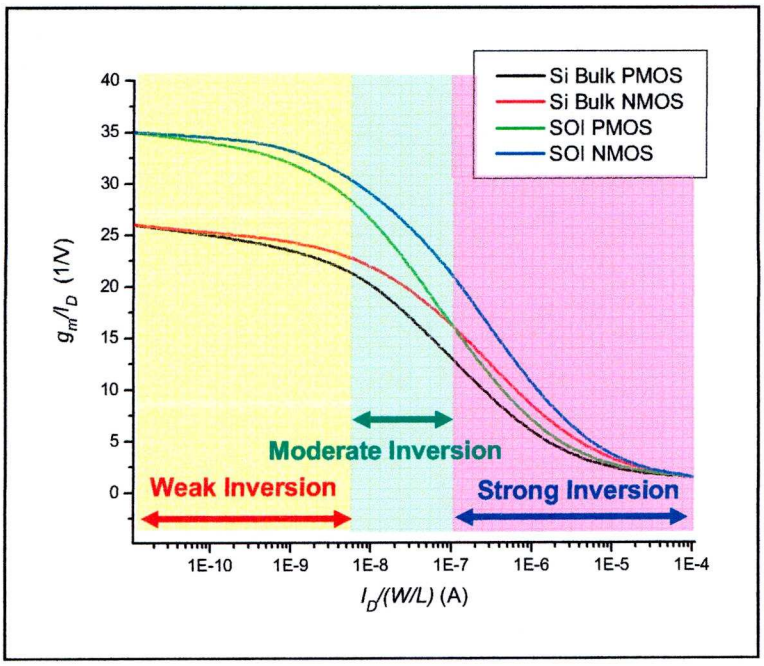

Figure 6.8, g_m/I_D vs $I_D/(W/L)$ curves for Bulk PMOS, NMOS and SOI PMOS, NMOS using EKV model [6.29].

The EPFL-EKV MOSFET model is a scalable and compact simulation model built on fundamental physical properties of the MOS structure [6.7]. This model is particularly useful in the design of low-voltage, low-current analogue, and mixed analogue-digital circuits using submicron CMOS technology. The model is also available in some SPICE packages which all together make it a highly desirable choice for the application that is discussed, designed and simulated in this thesis.

The EKV model gives,

$$g_m / I_D = \frac{1}{nU_T} \frac{1 - e^{(-\sqrt{IC})}}{\sqrt{IC}} \qquad (6.34)$$

$$IC = \frac{I_D}{2n\beta U_T^2}, \qquad \text{with } \beta = \mu C_{OX} \frac{W}{L}$$

where n is the linearized body factor, U_T is the thermal voltage, IC is the inversion coefficient, μ is the effective mobility, C_{ox} the gate oxide capacitance per unit area. IC determines the region of operation, $IC = 1$ at the weak-strong inversion transition, ($IC < 1$ w.i. & $IC > 1$ s.i.).

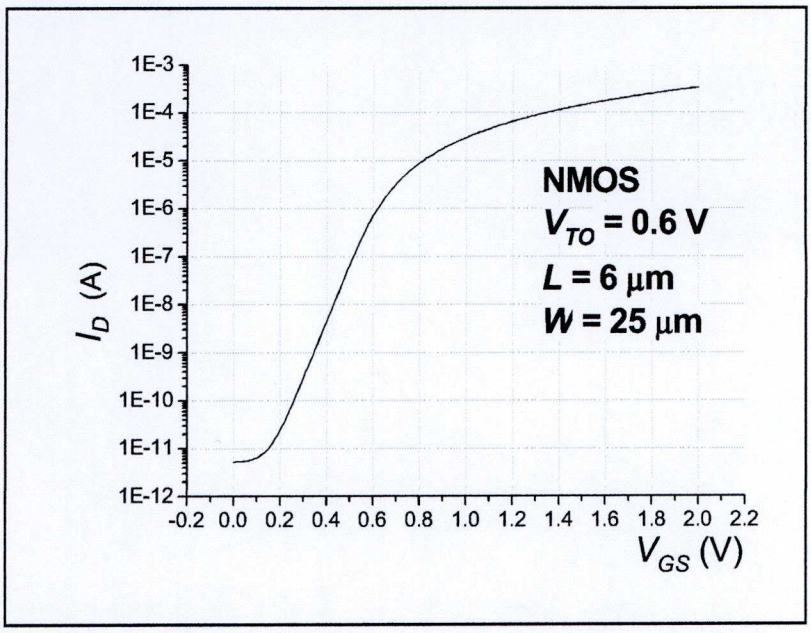

Figure 6.9, EKV model of drain current versus gate voltage for an Si bulk NMOS, $V_{TO} = 0.6$ V, $L = 6$ µm, $W = 35$ µm.

Equation (6.34) simplifies to standard weak and strong inversion transconductance equations.

In weak inversion, when $IC << 1$,

$$g_m = \frac{I_D}{nU_T} \tag{6.35}$$

In strong inversion, when $IC >> 1$,

$$g_m = \sqrt{2\beta I_D} \tag{6.36}$$

Although the EKV model was developed for Si bulk MOSFETs [6.7], it has been successfully extended to FD SOI MOSFETS. This has been possible because the gate-to-substrate coupling, which controls the subthreshold swing in weak inversion and the body factor in strong inversion, (i.e. 'n' factor discussed earlier), is given a single parameter 'n', as also presented previously in this thesis [6.29]. Figure 6.8 shows that good agreement has been achieved between measured and modelled g_m/I_D characteristics for both Si bulk and FD SOI.

6.6.9 *EKV noise*

The present version of the EKV model has "limited resources" for describing noise. Thus the modelling of noise is generally considered to be oversimplified compared to the theoretical and unified models given above and in particular does not support shot noise as discussed in 6.5. For our simplified simulations flicker noise is given by the simple empirical model,

$$\overline{I_{1/f}^{\,2}} = \frac{KF.g_m}{C_{ox}.f^{AF}}\Delta f \tag{6.37}$$

Where KF is the flicker noise coefficient, and AF is the flicker noise exponent. KF and AF are empirically derived according to the process, (AF = 1, KF = 1×10^{-27} for Si bulk NMOS, KF = 1×10^{-28} for Si bulk PMOS, KF = 7.5×10^{-27} for FDSOI NMOS and KF = 5×10^{-28} for FDSOI, PMOS).

The thermal noise is modelled as in (6.21) by,

$$\overline{i_{th}^{\,2}} = 4kT\frac{1}{R}\Delta f \tag{6.38}$$

Comparison of FD SOI versus Si bulk noise: Using the EKV noise model it can be seen that, as in section 6.5.2, if we consider the channel resistance as 3/2(1/ g_m) we again get (6.22),

$$\overline{i_{th}^{\,2}} = \frac{8}{3}kTg_m\Delta f \tag{6.39}$$

From (6.18) it is known $g_m \propto 1/n$. Therefore the reduction in 'n' improves the thermal noise in Si bulk over FDSOI.

For $1/f$ noise the carrier concentration increases in FD SOI due to confinement of the silicon layer which has been shown to increase $1/f$ noise in thin film FD SOI devices compared to Si bulk devices [6.38]. As

the body thickness of FD SOI devices decreases the $1/f$ noise will increase because of the enhanced carrier concentration.

However in many circuits, because the noise is referred back to the input, the reduction in the increase in (g_m/I_D) affects the transfer function for the noise back to the input and therefore, despite the device level increase in $(1/f)$ noise, even at low frequencies, there is often a reduction in the input equivalent noise for FD SOI compared to Si bulk circuits.

6.7 Conclusion

This chapter concentrated on the knowledge required for the design of specific hearing aid circuits and systems before we move onto the development and analysis of the circuits and systems in chapter 7.

At the end of chapter 5 it was concluded that the best hardware approach for hearing aid circuits and systems would be a hybrid analogue/digital approach with a VLSI technology. Chapter 6 therefore concentrates on researching technology to develop a hybrid frequency feature extraction system using the Lyon and Mead model of the cochlea. This research also concentrates on meeting the requirements of the specification discussed in chapter 4. Specifically we concentrate on low power, small silicon area and low noise circuits. Hence chapter 6 discusses low power analogue CMOS circuit design and identifies the need to minimise the ratio of the supply voltage and signal peak to peak voltage, as well as minimising the SNR. There are also practical limits to the amount the power consumption which be reduced. These limits are controlled by the intrinsic noise, limiting bias currents and reducing parasitic capacitance. Finally minimum power consumption for capacitive loads is shown to be achieved when g_m/I_D is maximised. Therefore chapter 6 continues by discussing the operation of CMOS transistors in weak inversion in order to maximise the g_m/I_D and operate with low voltage and low current. Next chapter 6 discussed the intrinsic noise within CMOS transistors because of its influence on minimising power consumption in analogue CMOS circuits.

Silicon–on–insulator (SOI) has been identified as a possible technology for hybrid CMOS hearing aid circuits and systems. Chapter 6 discusses the technology and the advantages and disadvantages of SOI technology compared to traditional Si bulk technology.

Finally chapter 6 looks at the ways in which SOI and Si bulk CMOS technologies can be modelled in all inversion region of transistor operation and including the influence of noise.

The knowledge from this chapter has then been used in the development and analysis of circuits designed in the next chapter.

References

[6.1] Lyon, R. F., C. Mead, C., "An analog electronic cochlea", IEEE Trans. Acoust., Speech, Signal Processing, vol. 36, pp. 1119-1134, Jul. 1988.

[6.2] Sarpeshkar, R., "Analog versus digital: extrapolating from electronics to neurobiology", Neural Computation, vol. 1, pp. 1601-1638, 1998.

[6.3] Welland, D., Phillip, S., Leung, K., Tuttle, T., Dupuie, S., Holberg, D., Jack, R., Sooch, N., Behrens, R., Anderson, K., Armstrong, A., et. al, "A Digital Read/Write Channel with EEPR4 Detection", Proc. IEEE Int. Solid-State Circuits Conf., pp. 276-277, Feb. 1994.

[6.4] Vittoz, E. A., "Future of analog in the VLSI environment", Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1372-1375, May 1990.

[6.5] Hogervorst, R., Huijsing, J. H., "Design of Low-Voltage, Low-Power Operational Amplifier Cells", Kluwer Academic Publishers, 1996.

[6.6] Tsividis, Y. P., Suyama, K., "MOSFET Modeling for Analogue CAD: Problems and Prospects", IEEE J. Solid-State Circuits, vol. 29, no. 3, pp. 210-216, 1994.

[6.7] Enz, C. C., Krummenacher, F., Vittoz, E. A., "An Analytical MOS Transistor Model Valid in all Regions of Operation and Dedicated to Low-Voltage and Low-Current Applications", Analogue integrated Circuits Signal Process., vol. 8, pp. 83-114, 1995.

[6.8] Allen, P. E., Holberg, D. R., "CMOS Analogue Circuit Design", 2nd ed., Oxford Uni. Press, 2002.

[6.9] Vittoz, E. A.,"Very low Power Circuit Design: Fundamentals and Limits", Proc. of ISCAS'93, pp. 1439-1442, Chicago, May 1993.

[6.10] Vittoz, E. A., "The Design of High-Performance Analogue Circuits on Digital Chips", IEEE J. Solid-state circ., vol. 20, no. 3, pp. 657-665, Jun. 1985.

[6.11] Sarpeshkar, R., Delbruck, T., Mead, C. A., "White Noise in MOS Transistors and Resistors," IEEE Circuits and Devices, vol. 9, no. 6, pp. 23-29, 1993.

[6.12] Wang, B., Hellums, J. R., Sodini, C. G., "MOSFET thermal noise modelling for analog integrated circuits", IEEE J. Solid-State Circuits, vol. 29, pp. 833-839, 1994.

[6.13] Tsividis, Y. P., "Operation and Modelling of the MOS Transistor", McGraw-Hill, New York, 1998.

[6.14] Buckingham, M. J., "Noise in Electronic Devices and Systems". Ellis Horwood Limited, Chichester, England, 1983.

[6.15] Arnaud, A., Galup-Montoro, C., "A compact model for flicker noise in MOS transistors for analogue circuit design", IEEE Trans. Electron Devices, vol. 50, pp. 1815-1818, Aug. 2003.

[6.16] Zafari, L., Jomaah, J., Ghibaudo, G., "Low frequency noise in multi-gate SOI CMOS devices", Proc. of EUROSOI 2006, pp. 141-142, Grenoble, Mar. 2006.

[6.17] Kwok K. Hung, Ping K. Ko, Chenming Hu, Yiu C. Cheng. "A physics-based MOSFET noise model for circuit simulators", IEEE Transactions on Electron Devices, vol. 37, no. 5, pp. 1323-1333, May 1990.

[6.18] McWhorter, A. L., "1/f noise and germanium surface properties", R. H. Kingston, editor, Semiconductor Surface Physics, University of Pennsylvania Press, pp. 207-228, Philadephia, 1957.

[6.19] Christensson, S., LundstrÄom, I., Svensson, C., "Low frequency noise in MOS transistors theory", Solid-State Electronics, vol. 11, pp. 797-812, 1968.

[6.20] Christensson, S., LundstrÄom, I., Svensson, C., "Low frequency noise in MOS transistors theory", Solid-State Electronics, vol. 11, pp. 813-820, 1968.

[6.21] Reimbold, G., "Modifed 1/f trapping noise theory and experiments in MOS transistors biased from weak to strong inversion influence of interface states", IEEE Transactions on Electron Devices, vol. 31, no. 9, pp. 1190-1198, Sept. 1984.

[6.22] Arnaud A., Galup-Montoro, C., "A compact model for flicker noise in MOS transistors for analogue circuit design", IEEE Trans. Electron Devices, vol. 50, pp. 1815-1818, Aug. 2003.

[6.23] Hooge, F. N., Vandamme, L. K. J., "Lattice scattering causes 1/f noise", Physics Letters, vol. 66A, no. 4, pp. 315-316, May 1978.

[6.24] Vandamme, L. K. J., de Werd, H. M. M., "1/f noise model for MOSTs biased in the nonohmic region", Solid-State Electronics, vol. 23, no. 4, pp 325-329, Apr. 1980.

[6.25] Hung, K. K., Ko, P. K., Hu, C., Cheng, Y. C., "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors", IEEE Trans. Electron Devices, vol. 37, no. 3, pp.654-665, Mar. 1990.

[6.26] Hung, K. K., Ko, P. K., Hu, C., Cheng, Y. C., "A physics-based MOSFET noise model for circuit simulators", IEEE Transactions on Electron Devices, vol. 37, no. 5, pp. 1323-1333, May 1990.

[6.27] Flandre, D., Gentinne, B., Eggermont, J. P., Jespers, P., "Design of Thin-film Fully-depleted SOI CMOS Analogue Circuits

significantly outperforming Bulk Implementations", Proc. of IEEE International SOI Conference, pp. 99-100, Oct. 1994.

[6.28] Veeraraghavan, S., Fossum J. G., "Short-channel effects in SOI MOSFET's", IEEE Trans. On Electron Devices, vol. 36, pp 522, 1989.

[6.29] Flandre, D., et al., "Fully-depleted SOI CMOS technology for heterogeneous. micropower, high-temperature or RF microsystems", Solid-State Electronics, vol. 45, no. 4, pp. 541-549, 2001.

[6.30] Flandre, D., "Silicon-on-insulator technology for high temperature metal oxide semiconductor devices and circuits", Mat. Sci. Eng. B, pp. 7-12, 1995.

[6.31] Renard, S., "Industrial MEMS on SOI", Journal of Micromechanics and Microengineering, Volume 10, no. 2, pp. 245-249, Jun. 2000.

[6.32] Mokwa,W., "Advanced Sensors and Microsystems on SOI", Int. Journal of High Speed Electronics and Systems, vol. 10, no. 1, pp. 147-153, 2000.

[6.33] Neumann, J. J., Gabriel, K. J., "CMOS-MEMS Membrane for Audio-Frequency Acoustic Actuation", Sensors and Actuators, A, vol. 95, pp. 175-182, 2002.

[6.34] Rombach, P., Mullenborn, M., Klein, U., Rasmussen, K., "The first low voltage, low noise differential silicon microphone, technology development and measurement results", Sensors and Actuators, A, vol. 95, pp. 196-201, 2002.

[6.35] Van Schaik, A., Shamma, S., "A neuromorphic sound localizer for a smart MEMS system", Analog Integrated Circuits and Signal Processing, vol. 39, pp. 267-273, 2004.

[6.36] Flandre, D., Vanhoenacker, D., "Thin-film fully-depleted SOI CMOS technology, devices and circuits for LVLP analogue/digital/microwave applications", Proc. of Semiconductor Conf, CAS '98, vol. 1, pp. 115-124, 1998.

[6.37] Flandre, D., Eggermont, J. P., De Ceuster, D., Jespers, P., "Comparison of SOI versus bulk performances of CMOS micropower single-stage OTAs", Electron. Letters, vol. 30 no. 23, pp. 1933-1934, 1994.

[6.38] Sadachika, N., Uetsuji, Y., Kitamaru, D., Weiss, L., Feldmann, U., Baba, S., Mattausch, H. J., Miura-Mattausch, M., "Fully-Depleted SOI-MOSFET Model for Circuit Simulation and Its Applications to 1/f Noise Analysis", Proc. of Simulation on Semiconductor Processes and Devices, 2004.

Chapter 7

7 ANALOGUE FILTER IMPLEMENTATION IN CMOS

7.1 Introduction

Following the discussion of CMOS technology design and modelling in chapter 6 the thesis can now discuss the development and analysis of a specific circuit for a hearing aid system. The development concentrates on a circuit for a frequency feature extraction system based on the Lyon & Mead model of the cochlea, as introduced in chapter 5.

The frequency feature extraction system requires the development of analogue CMOS filters. These filters require variable cut–off frequencies and variable gains. A cascade of these analogue filters can then be formed into a hybrid analogue/digital frequency feature extraction system which divides the input signal into different frequency taps which in turn can be processed by subsequent digital style circuits and systems.

In this chapter the filters are designed, modelled and developed using an operational transconductance amplifier, OTA. This design is used to highlight the differences between a FD SOI and Si bulk CMOS technology circuit implementation. The development of the circuits concentrates on improving the linearity of the filters to improve the dynamic range of the filters.

7.2 Operational Transconductance Amplifier

MOSFET devices are inherently transconductance devices therefore VLSI analogue filters are constructed from transconductance stages (g_m stages) in conjunction with capacitances. G_m-C filters, as they are known, have a distinct advantage over other possible analogue filter implementations because it is easy to adjust and externally control their characteristics through adjustment of the g_m stages [7.1]. The fundamental building block

of a VLSI analogue filter is the operational transconductance amplifier (OTA). Figure 7.1 shows an OTA circuit symbol and the ideal small signal equivalent circuit. The model (which considers ideal input and output impedances, equal to infinity) shows the OTA has a gain from the differential voltage input to the current output, (i.e. transconductance, g_m). A basic OTA consisting of just 6 transistors (4 NMOSFETs and 2 PMOSFETs) as shown in figure 7.2. The transconductance gain is well controlled and limited by the bias current, I_b making the device easy and simple to adjust through a current control.

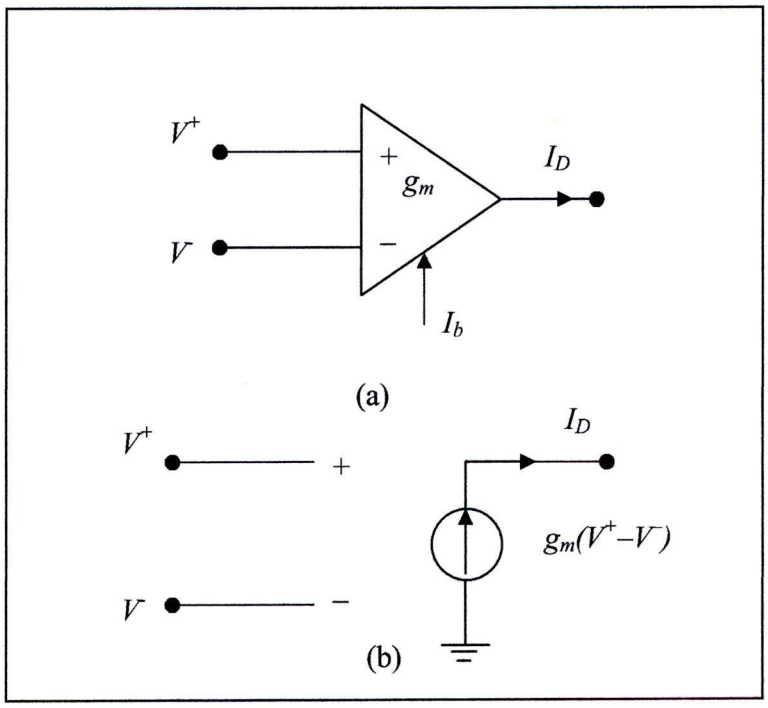

Figure 7.1, OTA (a) Symbol, (b) Ideal Small Signal Equivalent Circuit.

As highlighted in depth later, in weak inversion, the transconductance, g_m, is assumed to be proportional to the bias current I_b. Therefore,

$$g_m = hI_b \tag{7.1}$$

For MOSFETs functioning in weak inversion, the proportionality constant h is independent of the I_b, thus obtaining a linear dependence of the transconductance. However, h is dependant on temperature and the body factor, n, but independent of device dimensions.

The control of the bias current (I_b) is an essential part of the analogue filter circuit design. There are many techniques which are using MOSFETs for creating a current proportional to a given voltage; implementing one of them makes it easy to control the g_m, with an input voltage level. Figure 7.2 shows the basic circuit for a tuneable transconductance amplifier. Under quiescent conditions, $V_{id} = 0$ V, the two currents I_{d1} & I_{d2} add up to I_b,

$$I_b = I_{d1} + I_{d2} \tag{7.2}$$

The current mirror operates so that I_{D3}, which equals I_{D1}, is mirrored at I_{D4}. Hence $i_{out} = 0$ A, since $|I_{d1}| = |I_{d2}| = |I_{d3}| = |I_{d4}|$.

With a differential input V_{id} that causes $V_{GS1} > V_{GS2}$ then I_{D1} is increased by ΔI and I_{D2} is decreased by ΔI. The increase in I_{D1} causes a similar increase in I_{D3} which is mirrored at I_{D4}. In order to balance the current increase at I_{D4} and decrease at I_{D2}, then i_{out} becomes $2\Delta I$ and v_{out} is increased.

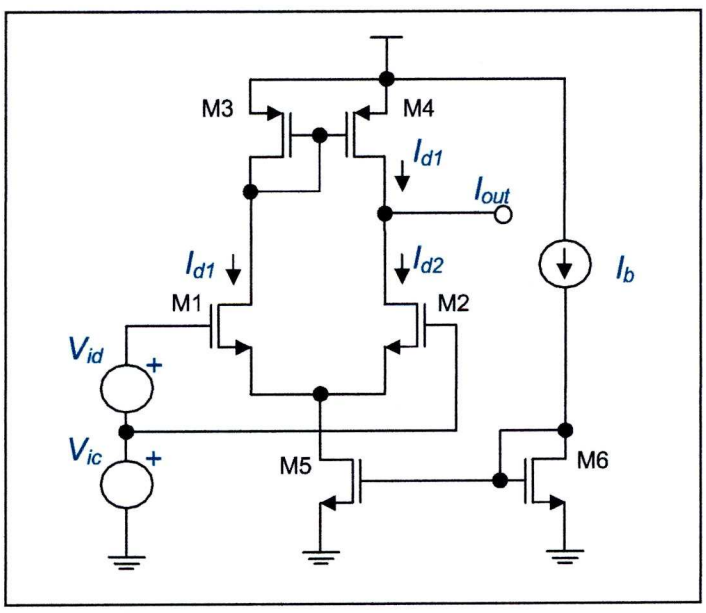

Figure 7.2, Tuneable Transconductance Amplifier CMOS circuit.

Figure 7.3 shows the DC transfer characteristics and transconductance of the OTA shown in figure 7.2 with tuneable g_m via variations in the bias current I_b. The simulation was carried out using 2 μm Si bulk devices

modelled by the EKV model in Hspice. The transfer characteristic of V_{id} to I_{out} is taken to be linear for a small region around $V_{id} = 0$ V. Taking this assumption, the compression of V_{id} to I_{out} is proportional to I_b.

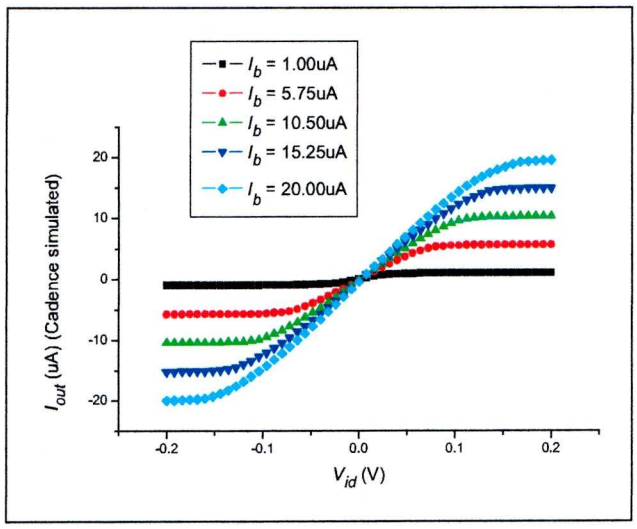

Figure 7.3, Simulation results for voltage to current compression controlled by bias current I_b.

7.3 Large Signal Analysis

For the transistors operating in weak inversion, the basic OTA has a transfer characteristic given by,

$$I_{out} = I_b \tanh\left[\frac{(v_{id+} - v_{id-})}{2nU_T}\right] \tag{7.3}$$

where n is the body factor and $U_T = kT/q$ is the thermal voltage.

7.3.1 Common mode range and minimum supply voltage

The constraining factor for the supply voltage is the subthreshold saturation drain-source voltage for each transistor and the gate-source voltages required for the maximum current in the diode-connected transistors. As discussed in chapter 6 the saturation drain-source voltage, V_{dssat}, for a subthreshold operating transistor is approximately 100 mV. The inputs must be contrained between $V_{ss} + V_b + (nV_{dssat})$ for the input to the ground side, (tail), of the OTA and $V_{dd} - V_{dssat}$ at the input to V_{dd}, (top),

of the OTA, where V_b is the voltage across the bias transistor M5. For a typical value of V_b = 0.8 V and V_{dd}, V_{ss} = 2.5 V, –2.5 V this gives a common mode input voltage level of between –1.6 V and 2.4 V.

7.4 Small Signal analysis

To carry out a small signal analysis we use the equivalent circuit shown in figure 7.4 and place a 10 pf capacitor at the output. This acts as an appropriate capacitive load C_L, and also simplifies the analysis. M5 and M6 shown in figure 7.2 were replaced with a simple current source. This simplification enabled Y to be taken as a virtual ground point because for small signal analysis the sum of the changes in I_{D1} and I_{D2} at node Y is zero. Figure 7.5 shows the simple small signal equivalent circuit and figure 7.6 shows the equivalent circuit with the load and parasitic capacitances included.

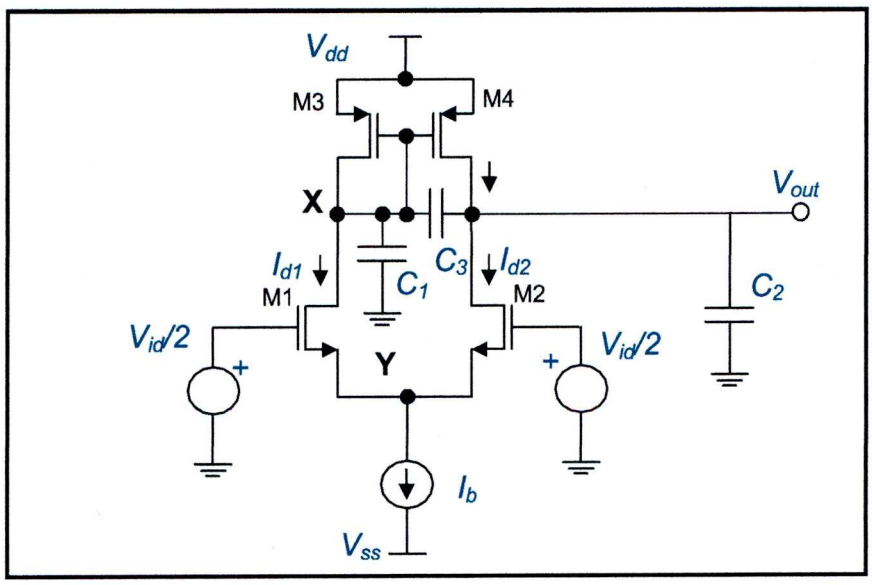

Figure 7.4, Parasitic Capacitances of OTA with differential input signal, (see (7.4)).

$$\begin{aligned} C_1 &= C_{gd1} + C_{db1} + C_{db3} + C_{gs3} + C_{gs4} \\ C_2 &= C_{gd2} + C_{db2} + C_{db4} + C_L \\ C_3 &= C_{gd4} \end{aligned} \tag{7.4}$$

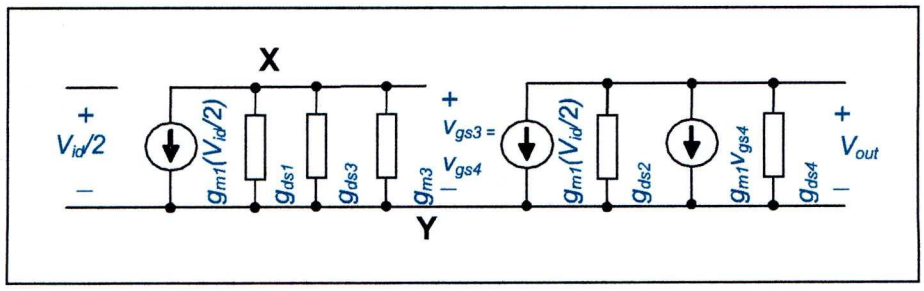

Figure 7.5, Small Signal equivalent circuit for an OTA.

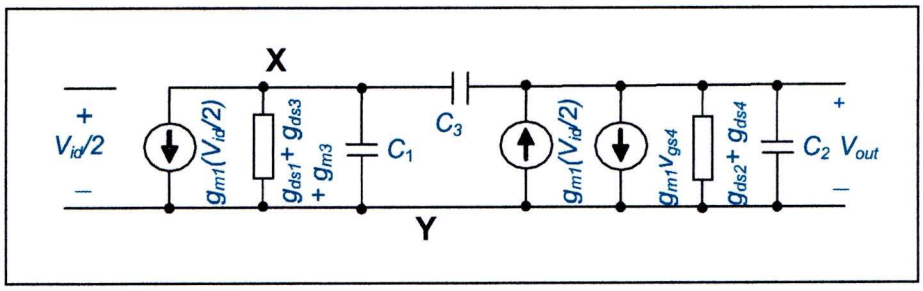

Figure 7.6, Small Signal equivalent circuit for an OTA including capacitances (see (7.4)).

From small signal analysis the current at Y must equal the bias current I_b, so a small increase in the current through M1, ΔI, caused by an increase in V_{id} is balanced by a reduction in the current through M2 by ΔI. Hence at the output the small signal change in current is $2\Delta I$. Thus the transconductance of the differential amplifier is given by,

$$g_{mOTA} = \frac{\partial I_{out}}{\partial V_{id}} = \frac{2\Delta I}{\Delta V_{id}} = \frac{\Delta I}{\Delta V_{id}/2} = \frac{\Delta I}{V_{GS1}} = g_{m1} \tag{7.5}$$

So the tranconductance of the differential amplifier is given by the transconductance of the simple transistor M1 or M2. Similarly from (6.18) the transconductance of the OTA can be calculated to be the same as the transconductance of a single transistor,

$$g_{mOTA} = \frac{\partial I_{out}}{\partial V_{id}} = \frac{I_b}{2nU_T} \tag{7.6}$$

and if $V_{id} = 0$V then $I_{d1} = I_{d2} = I_b/2$, therefore,

$$g_{mOTA} = \frac{I_{d1}}{nU_T} = g_{m1} \tag{7.7}$$

The output resistance of the differential amplifier is given from the drain source resistance of M2 and M4.

$$R_{out} = \frac{1}{g_{ds2} + g_{ds4}} \tag{7.8}$$

Therefore the open loop voltage gain is given by,

$$A_{vo} = \frac{v_{out}}{v_{in}} = \frac{i_{out} R_{out}}{v_{in}} = \frac{g_{m1}}{g_{ds2} + g_{ds4}} \tag{7.9}$$

Since g_{ds} is inversely proportional to L of the transistors, the gain is inversely proportional to the sum of the lengths of M2 & M4.

From figure 7.6 the transfer function can be derived as shown in appendix B, where the C_3 value can be ignored due to its negligible size.

$$A_v = \frac{g_{m1}}{g_{ds2} + g_{ds4}} \frac{\left(1 + \frac{sC_1}{(2g_{m3})}\right)}{\left(1 + \frac{sC_1}{g_{m3}}\right)\left(1 + \frac{sC_2}{g_{ds2} + g_{ds4}}\right)} \tag{7.10}$$

Appendix B also shows the poles and zeros are,

$$\begin{aligned} p_1 &= -\frac{g_{ds2} + g_{ds4}}{C_2} \cong -\frac{g_{ds2} + g_{ds4}}{C_L} \\ p_2 &= -\frac{g_{m3}}{C_1} \\ z &= --\frac{2g_{m3}}{C_1} \\ p_1 &< p_2 < z \end{aligned} \tag{7.11}$$

Thus the circuit has the dominant pole p_1 occurring at the output node and the zero is double the non-dominant pole p_2 in the middle of the current mirror.

Hence the circuit bandwidth can be approximated from the transition frequency f_t by;

$$f_t = \frac{1}{2\pi R_{out} C_L} \tag{7.12}$$

and the gain bandwidth product (GBW):

$$GBW \cong g_m R_{out} \frac{1}{2\pi R_{out} C_L} = \frac{g_m}{2\pi C_L} \tag{7.13}$$

7.5 FD SOI / Si bulk OTA Design, Fabrication and Testing

To enable comparison of Si bulk transistor technology OTA circuits and FD SOI transistor technology OTA circuits it was necessary to design and simulate physical layouts of the simple OTA shown in fig. 7.2. The circuits were designed and laid out using the Cadence® design software. Figure 7.7 shows a Cadence® design created for a typical Si bulk transistor technology OTA circuit. To link with the Cadence® designs the circuits were simulated with HSPICE using EKV models introduced in section 6.6.8, with 2 µm FD SOI transistors and for comparison, 0.5 µm Si bulk transistors.

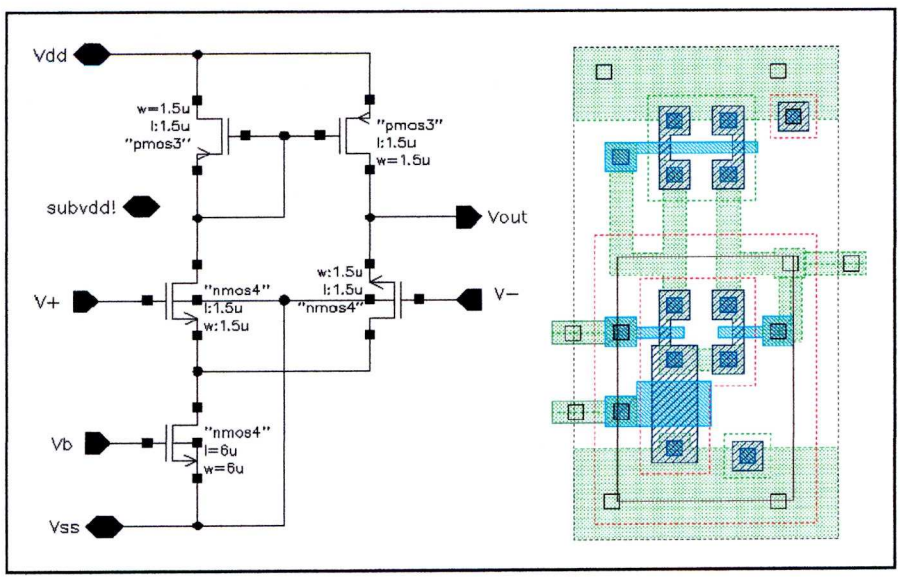

Figure 7.7, Cadence® design created for a typical Si bulk transistor technology OTA circuit.

For the comparison of Si bulk and FD SOI transistor technology OTA circuits the simple OTA shown in fig. 7.2 was used with the transistors sized to maintain minimum silicon area. The circuit was initially designed using the g_m/I_D methodology [7.2], with the specification of bandwidth greater than 20 kHz, low bias current I_b, (i.e. 25 nA and tuneable for a range approximately from 1 nA to 100 nA), and minimum supply voltage of ± 1 V. In the subthreshold region, the transconductance is independent of transistor dimensions; however to reduce noise and transistor dimensional mis-match, much larger than minimum transistor dimensions were chosen. Table 7-1 shows the transistor dimensions that were chosen.

Table 7-1, Parameters for basic OTA, (see figure 7.2).

Parameter	Value
W_{M1}, W_{M2}	100 µm
L_{M1}, L_{M2}	6 µm
W_{M3}, W_{M4}	25 µm
L_{M3}, L_{M4}	6 µm
W_{M5}, W_{M6}	25 µm
L_{M5}, L_{M6}	6 µm
I_b	25 nA
V_{dd}	1.5 V
V_{ss}	–1.5 V
V_{ic}	0 V

In order to justify the EKV model simulation results for FD SOI a physical FD SOI technology OTA was fabricated. The physical circuit was designed with the parameters in table 7-1. The design, fabrication and testing of a FD SOI transistor technology OTA circuit was one of the major achievements of this thesis. The fabrication was carried out by the author using the 2 µm FD SOI CMOS technology available at the UCL, Catholic University in Leuven-la-Neuve, Belgium; 'SmartCut' UNIBOND wafers (compatible with bulk Si CMOS processes), with a buried oxide layer of 400 nm, a Si substrate of 80 nm silicon, 31 nm for the gate oxide and a V_{TO} = 0.4 V were used [7.3]. Figure 7.8 shows a basic copy of the 2 µm FD SOI technology OTA circuit that was fabricated.

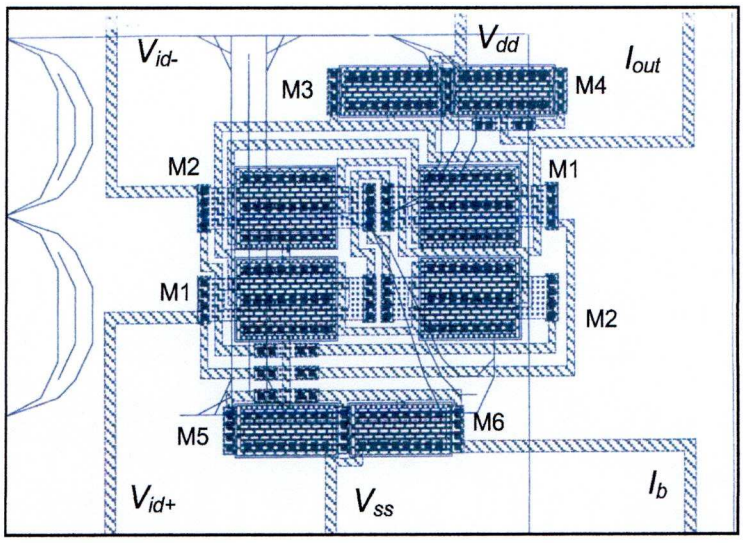

Figure 7.8, Basic copy of a 2 µm FD SOI technology OTA circuit fabricated at UCL, Catholic University in Leuven-la-Neuve, Belgium, based on the circuit shown in fig. 7.2 and parameter given in table 7.2

Figure 7.9 shows a comparison of simulated and experimental transconductance for FD SOI. The noise which appears for the experimental OTA1 is believed to be caused by a poor circuit contact whilst the offsets are a result of dimensional mismatch in the differential pair as a result of fabrication. Figure 7.9 provides a good validation of the EKV model for FD SOI in terms of the transconductance.

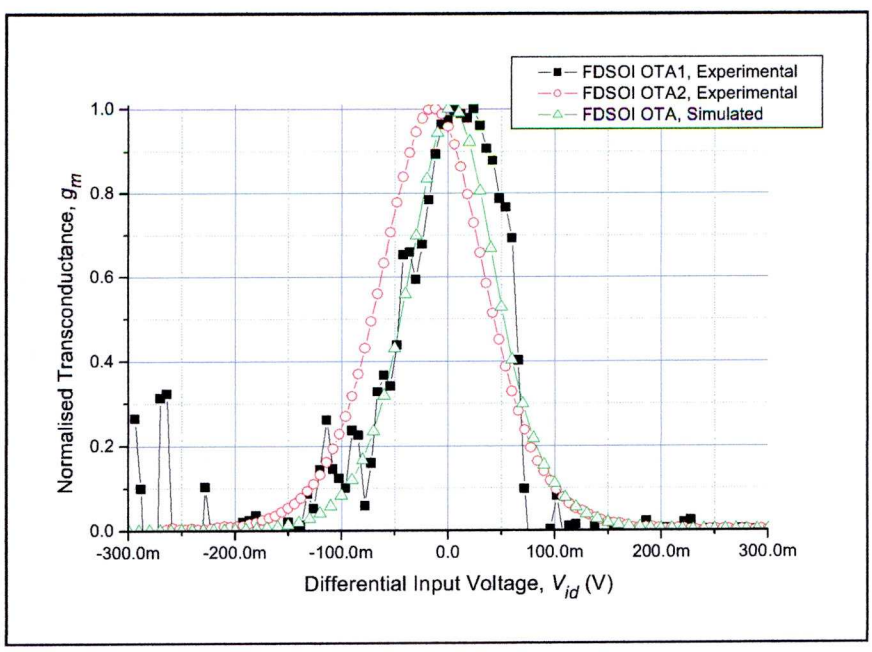

Figure 7.9, Simulated and experimental characteristics of the FD SOI OTA described in fig. 7.2 and table 7-1.

The increased value of g_m/I_D for FD SOI devices, discussed in section 6.6.2, can be seen in figure 7.10 to improve the bandwidth that can be achieved for a particular bias current. Hence it shows the power consumption required for a particular bandwidth will be reduced for a FD SOI OTA as compared to a SI bulk OTA.

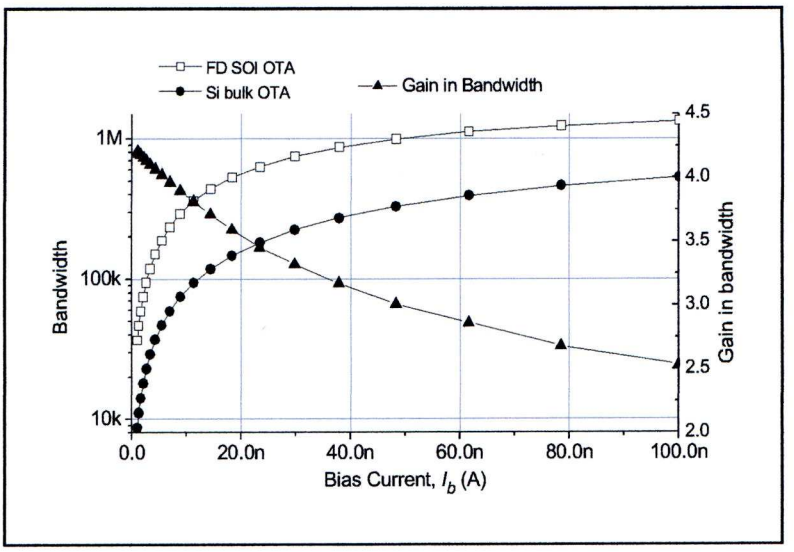

Figure 7.10, Comparison of bandwidth versus bias current, (as a measure of power dissipation), for an Si bulk and FD SOI basic OTA.

7.6 Noise equivalent Circuit for an OTA

For a single transistor the total output-noise current is found by summing each contribution to get,

$$i_{to}^{\ 2} = i_{th}^{\ 2} + i_{\not f}^{\ 2} \tag{7.14}$$

Since the equivalent output-noise current will be expressed in terms of an equivalent input-noise potential, we use,

$$i_{to}^{\ 2} = g_m^{\ 2} e_{eq}^{\ 2} \tag{7.15}$$

to give,

$$e_{eq}^{\ 2} = \frac{i_{th}^{\ 2}}{g_m^{\ 2}} + \frac{i_{\not f}^{\ 2}}{g_m^{\ 2}} \tag{7.16}$$

Using equations (6.39) and (6.40),

$$e_{eq} = \sqrt{\left(\frac{8}{3}kT\frac{1}{g_m}\right)+\left(\frac{KF.I_D{}^{AF}}{C_{ox}L^2 g_m{}^2 f}\right)} \qquad \left(\mathrm{V}/\sqrt{\mathrm{Hz}}\right) \qquad (7.17)$$

Figure 7.11a shows the basic OTA circuit with noise equivalent sources for each transistor. We can combine the noise sources of each transistor into an overall noise equivalent input for the OTA as shown in figure 7.11b to get,

$$e_{eqOTA}{}^2 = e_{n1}{}^2 + e_{n2}{}^2 + \left(\frac{g_{m3}}{g_{m1}}\right)^2 \left[e_{n3}{}^2 + e_{n4}{}^2\right] \qquad (7.18)$$

where we assume $g_{m1} = g_{m2}$ and $g_{m3} = g_{m4}$. We can also assume $e_{n1} = e_{n2}$ and $e_{n3} = e_{n4}$ so that,

$$e_{eqOTA}{}^2 = 2e_{n1}{}^2 + 2e_{n3}{}^2\left(\frac{g_{m3}}{g_{m1}}\right)^2 \qquad (7.19)$$

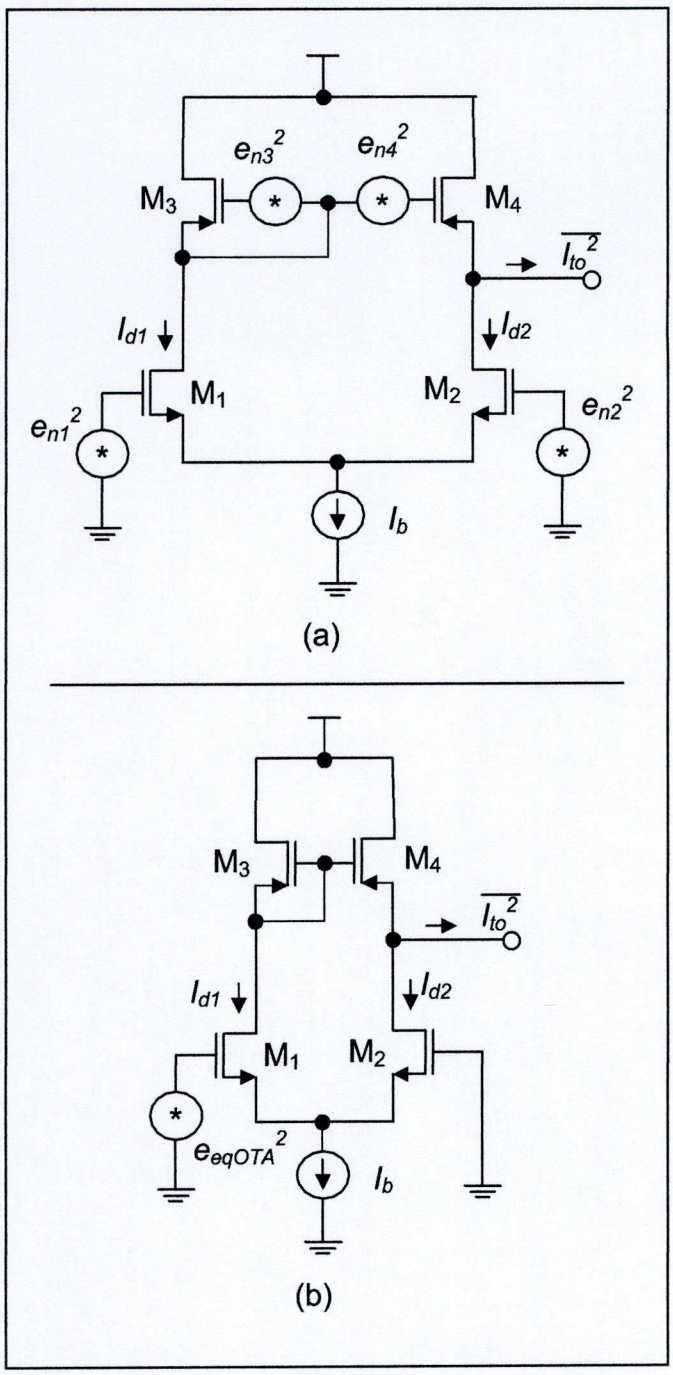

Figure 7.11, (a) Noise model of OTA with equivalent noise sources for each transistor. (b) Overall Equivalent noise model for OTA.

7.7 Linearity

The major limitation of OTAs is the restrictive differential input voltage swing around $V_{id} = 0$ V. The assumption that g_m remains constant limits the small signal input that can be taken before the OTA causes significant

distortion in the output signal. As a measure of the linear operating range the differential input voltage range that results in a maximum of 1% change in the transconductance is taken. The distortion at the output can be analysed by carrying out a Taylor expansion on the tanh relationship (7.3).

$$\frac{I_{out}}{I_b} = \tanh x = x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315} + ... \tag{7.20}$$

where,

$$x = \frac{V_{id}}{2nU_T} \tag{7.21}$$

Equation (7.20) shows that at $x = 1$ the output has cubic distortion of 33%, fifth harmonic distortion of 13% and seventh harmonic distortion of 5% whereas at $x = 0.5$ the output has cubic distortion of 4%, fifth harmonic distortion of 0.4% and seventh harmonic distortion of 0.0004%.

7.8 OTA in analogue filters

For the hearing aid structure presented in chapter 2 a feature extraction stage, modelled on the cochlear, typically requires a cascade of low pass filters [7.4].

7.8.1 First Order structure

First order filter structure can readily be built using OTAs. The simplest and typical first-order low pass G_m-C filter is shown in figure 7.12 and consists of an OTA with a closed loop connection and a single capacitor to create, at low frequencies a unity gain, but at high frequencies the circuit acts as a lossy integrator. The filter has an adjustable cut-off frequency (adjustable pole location) through the transconductance and/or capacitance, but does not have adjustable gain. The transfer function for the filter is,

$$\frac{V_{out}}{V_{in}} = \frac{1}{\frac{C}{g_m}s + 1} \tag{7.22}$$

The pole is at $s = -g_m/C$ determining the 3 dB cut-off frequency to be given by,

$$\omega_0 = \frac{g_m}{C} \tag{7.23}$$

Beyond the cut-off frequency the response declines at approximately 6 dB per octave.

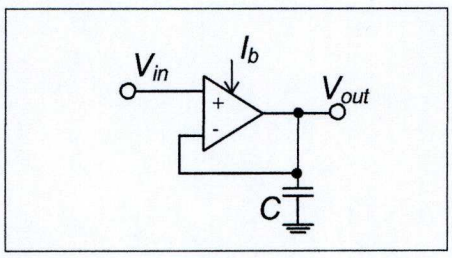

Figure 7.12, First-order low-pass filter, or follower integrator circuit using a single OTA. The bias current controls the cut-off frequency of the filter.

When designing a cascade of filter structures, it may be the case that the input impedance to each stage is not ideally infinite. If this is the case a unity gain buffer may be required. The input to figure 7.12 is assumed to be ideally infinite.

7.8.2 Second order filter structure

Second order filter structures are often more desirable because of the manner in which the filter characteristics can be adjusted. Second order filter structures also find direct application in the design of higher order filter structures. A simple second order structure using a cascade of two follower integrators shown in figure 7.13.

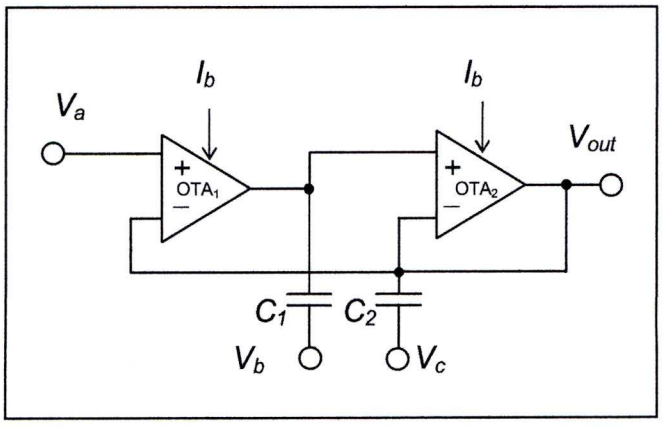

Figure 7.13, Second order G_m-C filter using two OTAs.

Table 7-2, Transfer functions for Second order G_m-C filter structure shown in figure 7.13.

Circuit Type	Input Condition	Transfer function	ω_0 Resonant Frequency	Q Quality Factor
Lowpass	$V_{in} = V_a$ $V_b = V_c = \text{Gnd}$	$\frac{g_m^2}{s^2C_1C_2 + sC_1g_m + g_m^2}$	$\frac{g_m}{\sqrt{C_1C_2}}$	$\sqrt{\frac{C_2}{C_1}}$
Bandpass	$V_{in} = V_b$ $V_a = V_c = \text{Gnd}$	$\frac{sC_1g_m^2}{s^2C_1C_2 + sC_1g_m + g_m^2}$	$\frac{g_m}{\sqrt{C_1C_2}}$	$\sqrt{\frac{C_2}{C_1}}$
Highpass	$V_{in} = V_c$ $V_b = V_a = \text{Gnd}$	$\frac{s^2C_1C_2}{s^2C_1C_2 + sC_1g_m + g_m^2}$	$\frac{g_m}{\sqrt{C_1C_2}}$	$\sqrt{\frac{C_2}{C_1}}$

The structure is very useful because by adjusting the specific excitations at V_a, V_b or V_c the transfer response can be either low-pass, high-pass or bandpass as listed in table 7-2. By using $g_{m1} = g_{m2} = g_m$ the circuits behave as adjustable ω_0 circuits with a contants Q-factor given by fixed poles (i.e. fixed along the jw axis) and set by the capacitor ratio, $Q = \sqrt{(C_2/C_1)}$.

As discussed early in chapter 2 the outer hairs cells of the cochlear have an active effect and actually add energy to the system by providing active

gain. In terms of a hearing aid active and adjustable gain is highly desirable for hearing compensation and compression and could also be used as an important feature for noise cancellation. In terms of an analogue filter cascade modelling the cochlear, it is necessary to have second-order filters with an adjustable Q-factor so that an underdamped response can provide a resonant peak as illustrated in figure 7.14. The combination of small resonant peaks from individual filter channels in the cascade produces a wider large composite "pseudoresonance" which more closely matches the natural cochlear than a single underdamped resonant gain response of a signal channel.

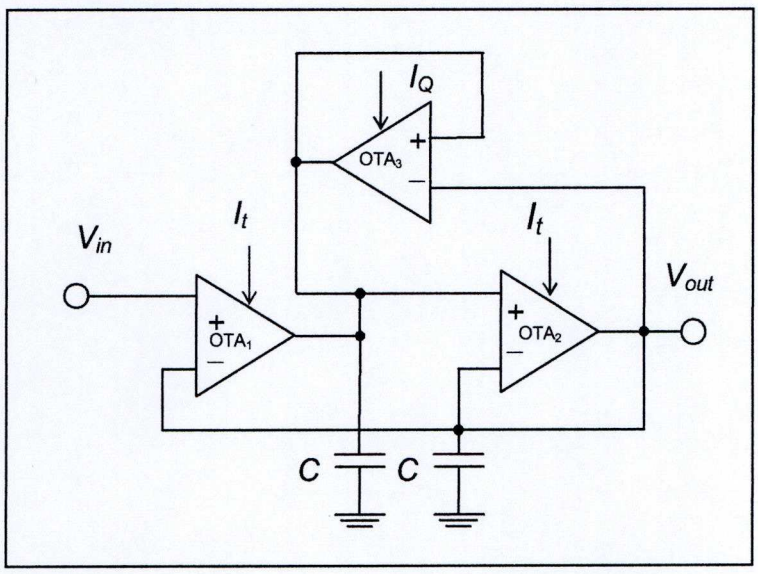

Figure 7.14, Second order low pass filter structure with controllable Q.

A second order filter that has been widely used for analogue filters is shown in figure 7.14 and has a transfer function

$$H(s) = \frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{1 + \varpi/Q + \varpi^2} \tag{7.24}$$

where $\tau = C/g_t$; $Q = \dfrac{1}{2(1-\alpha_t)}$ and $\alpha_t = g_Q/(2g_t)$,

and where $g_t = g_{mOTA1} = g_{mOTA2}$ and $g_Q = g_{mOTA3}$.

This is based on the previous second order filter containing two cascaded follower integrators with the addition of another amplifier. The third amplifier creates an underdamped response (about gain) which can be used to better model the active operation of the cochlear [7.4]. The third amplifier allows the Q factor to be controlled by the transconductance of this amplifier.

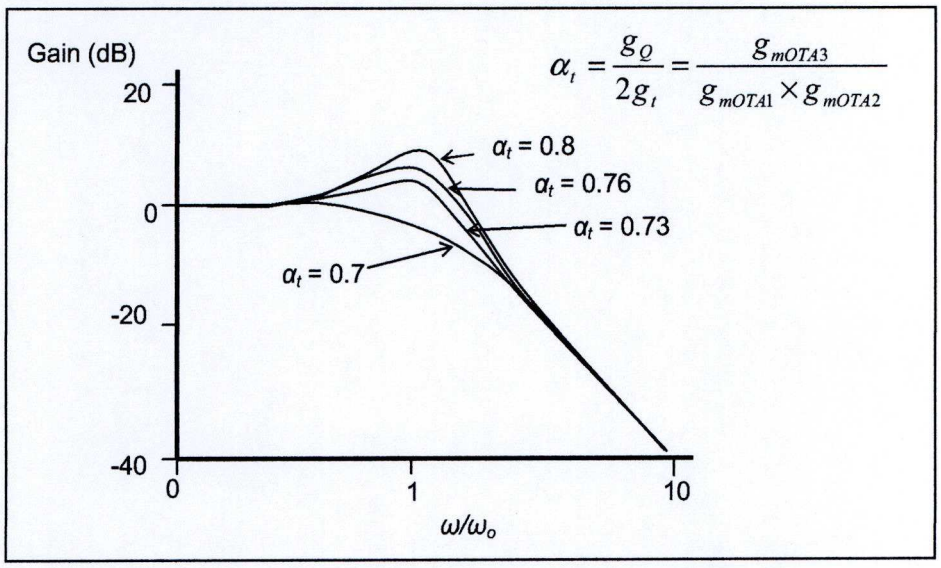

Figure 7.15, Gain magnitude response for second order low pass filter, with controllable Q, figure 7.14, for various values of α_t.

If Q is less than 0.707 the response has a purely low-pass characteristic where as if Q is greater then 0.707 the response is underdamped with a resonant peak. The height of this peak increases and its width decreases as Q further increases until the system becomes unstable.

Equation (7.24) has two criteria for the stability of small signals, when $Q = \infty$, i.e. $\alpha_t = 1$, and for large signals [7.5]. The large signal stability limit has been derived using a piecewise linear approach [7.6] which corresponds to $Q = 2.63$ or $\alpha_t = 0.809$. The stability is summarized as,

$0 < \alpha_t < 0.809$	unconditionally stable.
$0.809 < \alpha_t < 1$	small-signal stable, large signal unstable.
$\alpha_t > 1$	unstable.

Since it would not be possible to guarantee the input would remain a small signal it seems prudent to ensure stability for large and small signals.

In order to achieve a high Q in the second order filter structure as shown in figure 7.14 it is required that g_Q is almost twice as large as g_t. However increasing the transconductance of the feedback amplifier results in an increase in the bias current required for the amplifier, I_{bQ}, from equ. (6.18). This increase in bias current can cause problems because it can cause the filter to become large-scale unstable when $\alpha_t > 0.809$ or from (7.24 & 6.18) when $I_{bQ} > 1.62I_b$. This instability can be avoided by increasing the linear range of the forward feeding amplifier more than the feedback amplifier. It is apparent that the ratio of feed forward amplifier linear range to feedback amplifier linear range must be at least 2/1.62 = 1.23 [7.5].

7.8.3 Dynamic range

The fundamental goal of designing an appropriate analogue filter for hearing aid application comes down to creating a low power filter with a large enough dynamic range.

When we talk about dynamic range we fundamentally want our circuit to be able to operate successfully with input signals with as large an amplitude as possible and also to be able to successfully transform as small a signal as possible. To this end the upper limit of the dynamic range is controlled by the amplitude of the input signal that does not cause too much distortion in the output signal and the lower limit is controlled by the intrinsic noise of the circuit. That is, the input signal must be suitably larger than the inferred equivalent noise input that appears at the input so that the output signal can easily be distinguished between what is actually a transformed input signal and what is simply noise created by the devices in the circuits by their very nature.

In circuits such as the filter structure using OTAs, as discussed above, the upper limit of dynamic range is controlled by the linearity of the OTAs [7.7]. This is because the OTA is for large signals, a non-linear, device which is taken to be linear for small signal inputs. Hence if the input

signal becomes too large the OTA moves away from the pseudo-linear operation to an increasing non-linear operation which results in the distortion of output signals. For ease of development of the OTA we assumed that better linearity directly equates to an improvement in the upper limit of the dynamic range of a filter structure. This is broadly true; however, as discussed later, the linearity of specific OTAs within a filter structure effect the operation of the filter in different ways.

7.8.4 Linearity

As mentioned above in order to develop the analogue filter design the basic challenge was to improve the linearity of OTA building blocks within the filter structures. This is an idea that has been used to develop this type of cochlear filter before. However we are introducing the OTA building blocks built with FD SOI devices for the reasons discussed in chapter 6. Therefore the new challenge has been to develop and improve the linearity of an FD SOI OTA circuit.

As device sizes, supply voltages and power consumption have been continuously scaled down the challenge of keeping linearity with a reasonable input signal level has been recognised as having particular importance. We have previously discussed that operating the OTA in weak inversion has been shown to have particular advantages for supply voltage and power consumption scaling. Several circuit techniques have been proposed in literature to improve the linearity of subthreshold CMOS OTAs. These techniques include using cross coupled differential pairs [7.8], adaptive biasing [7.9], source degradation [7.10] and bump linearization [7.11]. All these techniques aim to improve the linearity of the subthreshold 'tanh' expression previously given in (7.3). There are two ways that all these techniques basically achieve improved linearity of the OTA. In order to produce a reasonable linear conversion over as wide an input voltage range as possible g_m must be approximately constant for as wide a range as possible. Firstly an improvement in linearity can be achieved by eliminating non-linearities in the voltage-to-current characteristics which cause g_m to be non-constant. Secondly a reduced

value of g_m results in the variation of I_{out} occurring over a wider range of V_{id} and thus there is a better linear approximation.

Figure 7.16 shows an OTA with two asymmetric differential pairs. Each pair of transistors is unequally sized by the same relative width-to-length ratio (m), so transistors M_1 and M_2 are a ratio of m scaled size of transistors to M_m. The two sets of transistors are then connected to form a cross-coupled differential pair. By having an asymmetric pair that have an equal but opposite size ratio the maximum transconductances for each pair have offset voltages around $V_{id} = 0$ V that are equal but opposite in magnitude as can be seen in figure 7.16. Taking the overall transocnductance of both transistor pairs equates to the sum of the two offset transconductances at a particular V_{id}. By selecting an appropriate ratio for the differential pairs the overall transconductance has a flattened peak around $V_{id} = 0$ V. Using the first derivative to give the transconductance and setting the second derivative of g_m to zero we can find the maximum linearity is given by $m = 2 + \sqrt{3}$. The technique has been shown to increase the linearity approximately 4 fold [7.8].

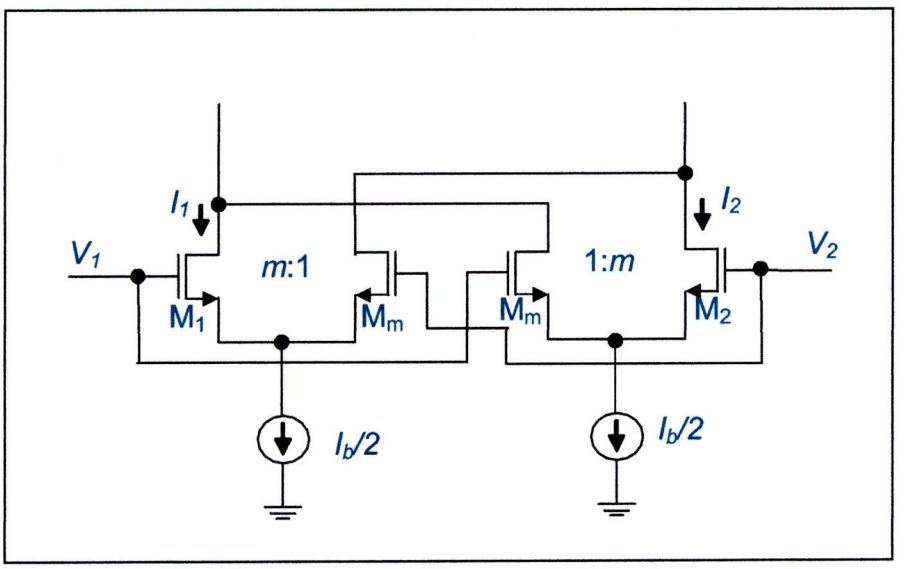

Figure 7.16, Asymmetric differential pairs with common ratio m:1, for the linearization of an OTA.

Source degradation: Figure 7.17 shows four different techniques that all use source degradation to improve linearity. The idea of source

degradation (SD) is to reduce the current flowing through the differential pair, therefore reducing the transconductance and hence increasing linearity. The simplest and most widely known technique is to include a resistor between the two arms of a differential pair as shown in figure 7.17a. This technique effectively reduces current in the differential pair M_1 & M_2 by reducing the differential input voltage in equation (6.18) by $I_{out}R_{SD}$. This then reduces the OTA transconductance g_m so that it is given by,

$$g'_m = \frac{g_m}{1 + g_m R_{SD}} \tag{7.25}$$

where g_m is the transconductane of the differential pair M_1 & M_2 (M_1 & M_2 being matched).

From equation (7.25) when $R_{SD} >> 1/g_m$ then g_m is significantly reduced and high linearity can thus be obtained. The disadvantage of this technique is the large resistor value required and the restriction the technique places on the size of g_m to achieve high linearity. In order for this technique to achieve a higher g_m value, g_m must be increased which in turn requires a higher power consumption.

Figure 7.17b shows the second technique which also uses source degradation with a transistor to act as a single diffuser [7.9] rather than a resistor. The same technique has been described for above threshold operation in [7.9] and [7.8] for subthreshold. The conductivity of the diffuser is set by the width-to-length ratio compared M_1 & M_2, m, and by the V_{GS} of the transistor. The best linearity is achieved by $m = 0.25$. This results in an eight fold increase in linearity; however the common-mode input voltage to the circuit must be around V_{GS}.

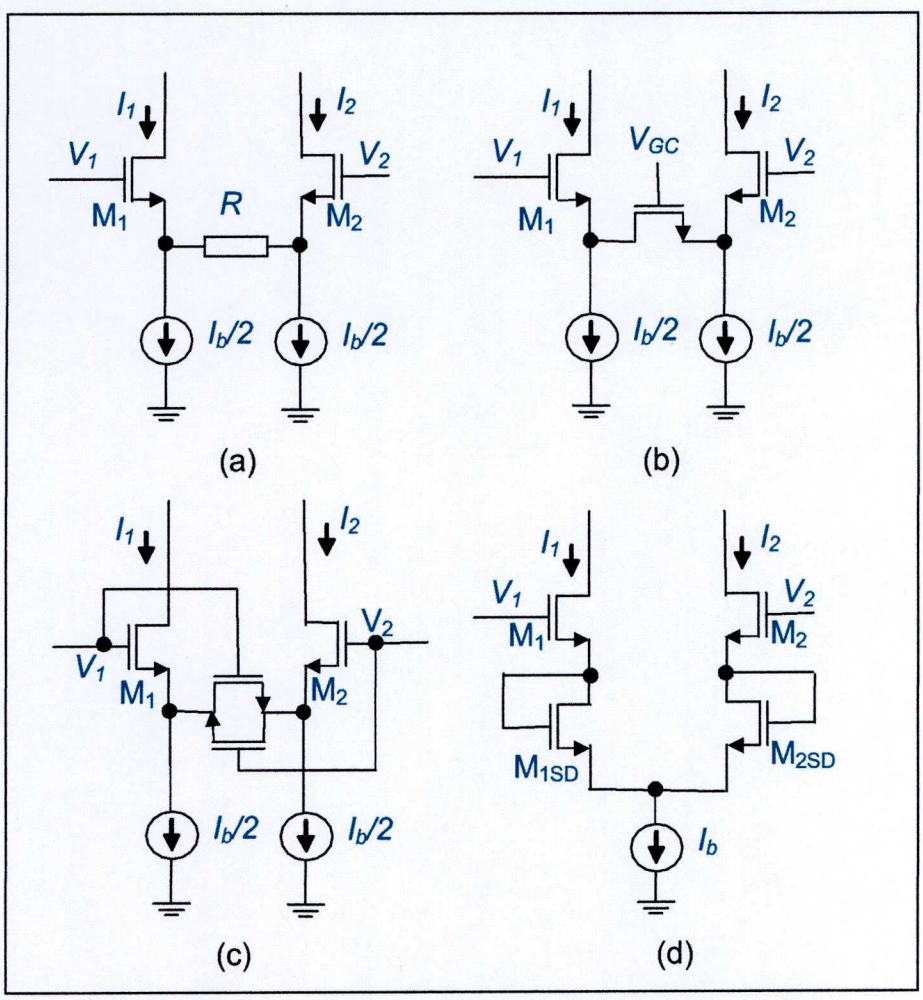

Figure 7.17, Differential pair and four different source degradation techniques for the linearization of an OTA; a) Single resistor diffuser; b) single transistor diffuser; c) symmetric transistor diffusers; d) diode connected series transistors.

The third technique uses a symmetric pair of diffusers, again with a width-to-length ratio compared to M_1 & M_2, of m, as shown in figure 7.17c. The maximum linearity is achieved when $m = 0.5$ [7.8]. This technique has the advantage over the simple resistor that when the amplitude of either input signal rises the corresponding diffusion transistor becomes more biased and synthesised resistance is reduced. This allows the drop in g_m caused by the source degradation resistance to be reduced and hence a higher g_m is possible. This technique has been shown to increase the linearity by approximately four times [7.8].

Figure 7.17d shows the OTA differential pair including a diode connected source degradation transistor in each arm. The diode connected source

degradation transistor, $M1_{SD}$ & $M2_{SD}$, increases the voltage at the source of the input differential pair transistors which therefore decreases the drain current. The source degradation diode connected transistor has the effect of reducing the tranconductance by increasing the effective body effect coefficient, n_e, so that in (6.18) if $n = n_e$ it becomes,

$$n_e = \begin{cases} n_0 & \text{(no SD diodes)} \\ \left(n_0 + n_0^{\,2}\right) & \text{(SD diodes)} \end{cases} \tag{7.26}$$

where n_0 is the body effect coefficient for the technology. For typical n values we expect the linear range for Si bulk to be increased by a factor of 2.5, whilst for FD SOI the linear range will be increased by a factor of 2.1. The linearity can theoretically be further increased by adding a further diode connected transistor in series thereby further decreasing the drain current of the differential pair for a certain V_{id} input. This increase in linearity has to be balanced with an increase in the minimum supply voltage and decrease in the common-mode operating range because the addition of the source degradation transistors adds an additional series voltage drop between the differential pair and the bias current transistor along with the relative increase in the body effect coefficient. There is also an increase in thermal intrinsic noise. For the simple OTA shown in figure 7.2 the typical common mode input voltage range was –0.6 V to 1.4 V. Table 7-3 shows how an increasing number of source degradation diodes affects the common mode input voltage range and shows there is an effective limit of two SD diodes in series for a supply voltage of ± 1.5 V and a limit of one SD diode for ± 1 V.

Table 7-3, Effect of differing numbers of Source Degradation transistors on Common Mode input Voltage and the effect on linearity.

Number of SD diodes	Approximate Factor Increase in linear range (Si bulk, FD SOI)	Common mode input voltage range
0	1, 1	–0.6 V to 1.4 V
1	2.5, 2.1	–0.35 V to 1.4 V
2	4.1, 3.5	–0.06 V to 1.4 V
3	–	1.64 V to 1.4 V (i.e. not possible)

Bump linearization: The next technique for improving the linearity is bump linearization. The bump linearization technique widens the linear range by eliminating some of the non-linearity in the tanh function given by (7.3). The technique uses the addition of a simple current correlator to the circuit. This comprises of the addition of two transistors connected in series to form a central arm between the two arms of the differential pair. With the pair of middle transistors connected to the differential *+ve* and *– ve* voltage input the current through the transistors has a bump-shaped function. By controlling the *W/L* ratio of the central 'bump' transistors we can control the characteristic shape of the 'bump' current. Finally by connecting the bump transistors to the current source I_b the bump transistor path takes a 'bump' current away from the 'tanh'-like current function of the differential pair outer arm current paths. Using the correct *W/L* ratio the bump current acts to flatten and linearize the region of the differential output current function around the $V_{id} = 0$ V.

The bump transistors have no detrimental effect on the power consumption or supply voltage required since they only add two transistors in series, which is equivalent to the transistors in the outer arms, and all the current in the outer arms and central bump transistor sums to I_b.

Taking S to be the parameter of the ratio of the outer and middle transistor dimensions we can compute the bump current to be given by,

$$I_{mid} = \frac{I_b}{1 + \frac{4}{S}\cosh^2 \frac{V_{id}}{2nU_T}} \tag{7.27}$$

$$S = \frac{W/L_{bump}}{W/L_{diff.pair}}. \tag{7.28}$$

From equation (7.27) the effect of the bump transistors on the overall OTA output current can be computed to be given by,

$$I_{out} = \frac{I_b \tanh\left(\frac{V_{id}}{2nU_T}\right)}{1+\frac{S}{4}\operatorname{sech}^2\left(\frac{V_{id}}{2nU_T}\right)} \tag{7.29}$$

For $S = 0$ we can see this reduces to (7.3). Delbrück [7.11] gives details that show the value we need for the maximally linear response is produced when the value of $S = 2$.

We can see the increase in the linear range by doing a Taylor expansion of (7.3) and (7.29) centred around $V_{id} = 0$ V, using $S = 2$ as the optimum linear case.

$$\frac{I_{out}}{I_b} = \tanh x = x - \frac{x^3}{3} + \frac{2x^5}{15} - \frac{17x^7}{315} + ... \tag{7.30}$$

$$\frac{I_{out}}{I_b} = \frac{2\tanh x}{2+\operatorname{sec}h^2 x} = \frac{2x}{3} - \frac{8x^5}{135} + ... \tag{7.31}$$

where,

$$x = \frac{V_{id}}{2nU_T} \tag{7.32}$$

The transconductance is reduced by 1/3 and the tanh function has a 3rd harmonic distortion of approximately 8% compared to no 3rd harmonic distortion in equation (7.30), the linearized function. Also the linearized function has a 5th harmonic distortion which is much less than the tanh function.

7.8.5 Linearized FD SOI tuneable OTA

In order to produce a reasonable linear conversion over as wide an input voltage range as possible two linearization techniques have been shown to be effective for an FD SOI OTA circuit [7.12] and provide up to a ten fold increase in linearity. These techniques are source degradation and bump

linearization. Figure 7.18 shows the linearized OTA that has a source degradation diode connected transistor in each arm, ($M1_{SD}$, $M2_{SD}$) and a pair of bump series connected transistors between the two arms, ($M1_{bump}$, $M2_{bump}$). The bump transistors' gates are connected to the gates of the source degradation diodes, therefore the $(W_{bump}/L_{bump})/(W_{SD}/L_{SD})$ ratio controls the bump linearization and allows the smallest possible bump transistors to be used. This ratio equals 2 to provide the maximum linearization.

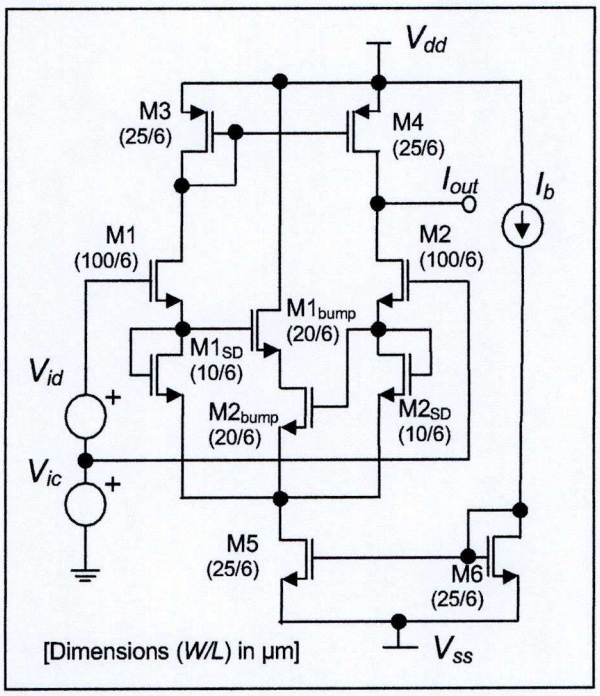

Figure 7.18, Linearized FD SOI tuneable OTA.

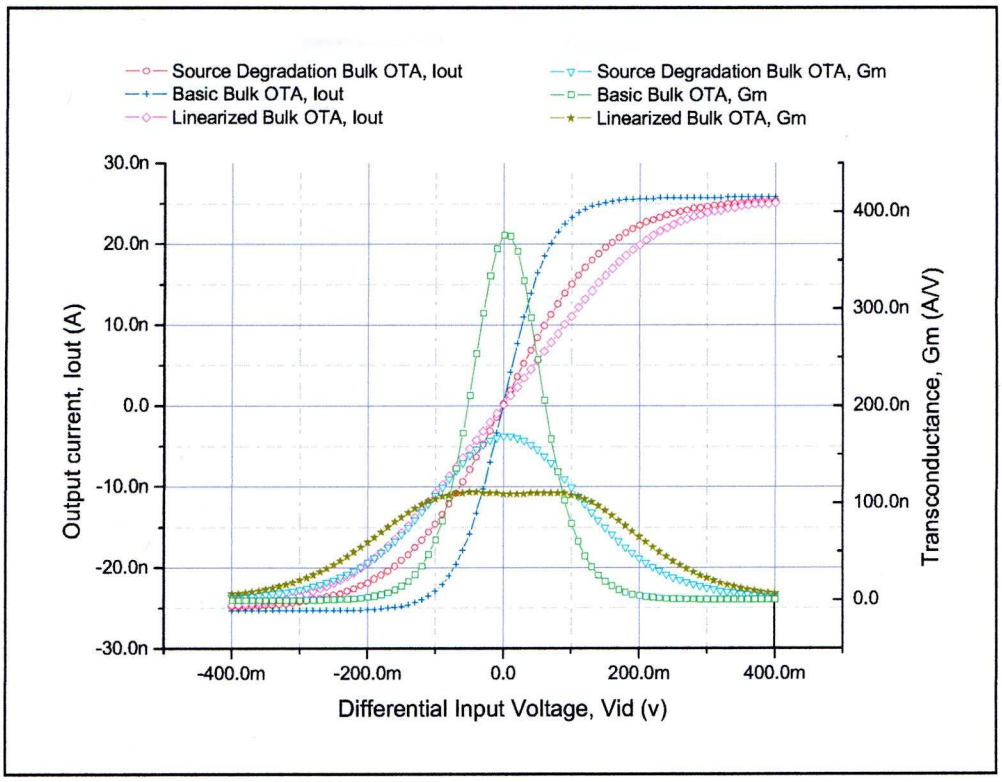

Figure 7.19, Voltage-to-current characteristics and transconductance of a Si bulk OTA with no linearization; Source degradation linearization and Source degradation and Bump Linearization.

The circuits including the linearization techniques described above were simulated with HSPICE using EKV models introduced in section 6.6.8, with 0.5 μm Si bulk transistors. Figure 7.19 shows how source degradation reduces the transconductance for a given V_{id} and therefore reduces the rate of change of I_{out} around V_{id} = 0 V, therefore improving linearity. The figure also shows how the addition of source degradation and bump linearization reduces the transconductance and also makes the transconductance closer to a constant value around V_{id} = 0 V. Again this results in the I_{out} verus V_{id} becoming closer to linear around V_{id} = 0 V.

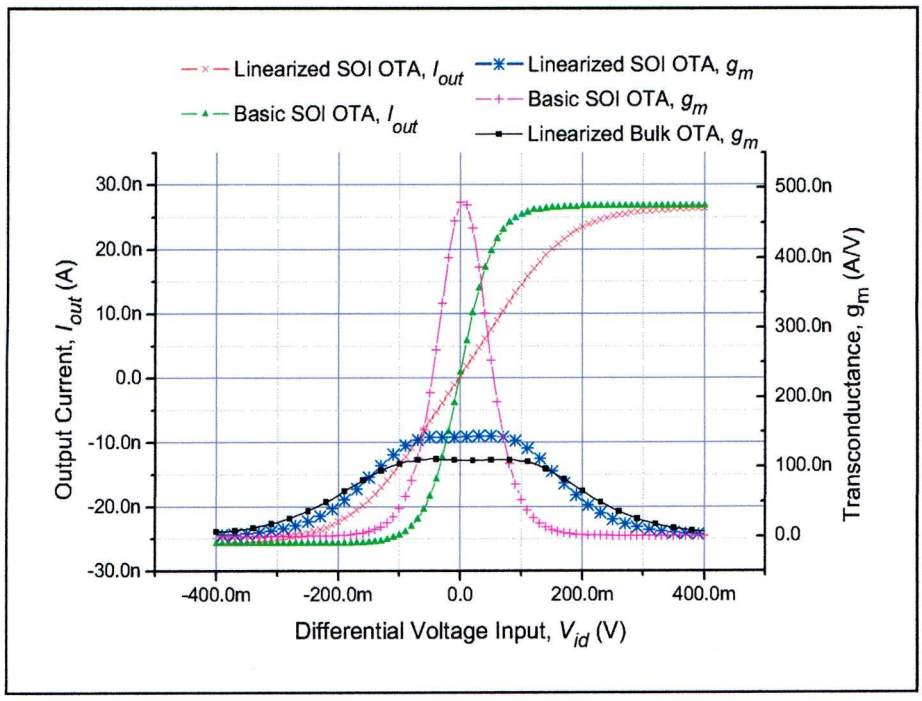

Figure 7.20, Voltage-to-current characteristics and transconductance of the basic and linearized OTA in Si bulk and FD SOI.

Figure 7.20 shows a comparison of FD SOI and SI bulk OTAs and how the linearity for each is improved. The body effect coefficient increases g_m for the FD SOI OTA and these results can still be seen comparing the linearized OTAs.

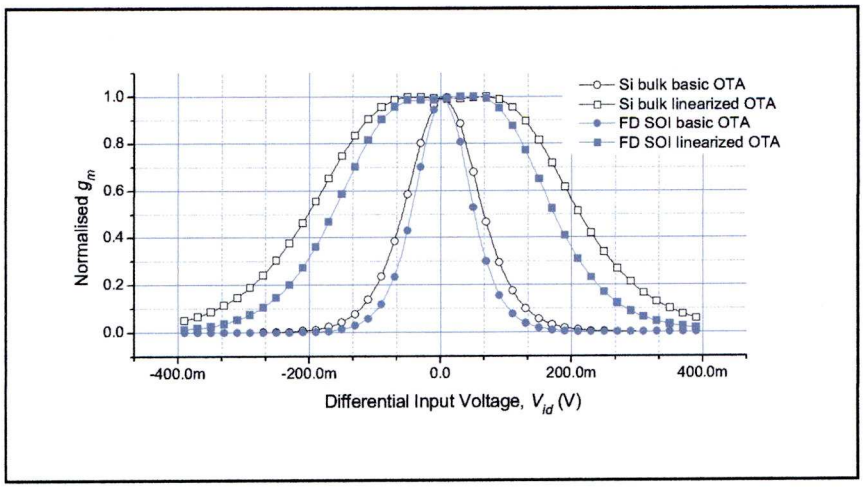

Figure 7.21, Normalized tranconductance for Si bulk and FD SOI OTA circuits.

The improvement in linearity was measured using the linear approximation as 1% of the normalised transconductane and provides a measure for the improvement in linearity, table 7-4. Fig. 7.21 shows how the OTA linear

differential voltage input range is improved for the normalised transconductance and how FD SOI and Si bulk compare.

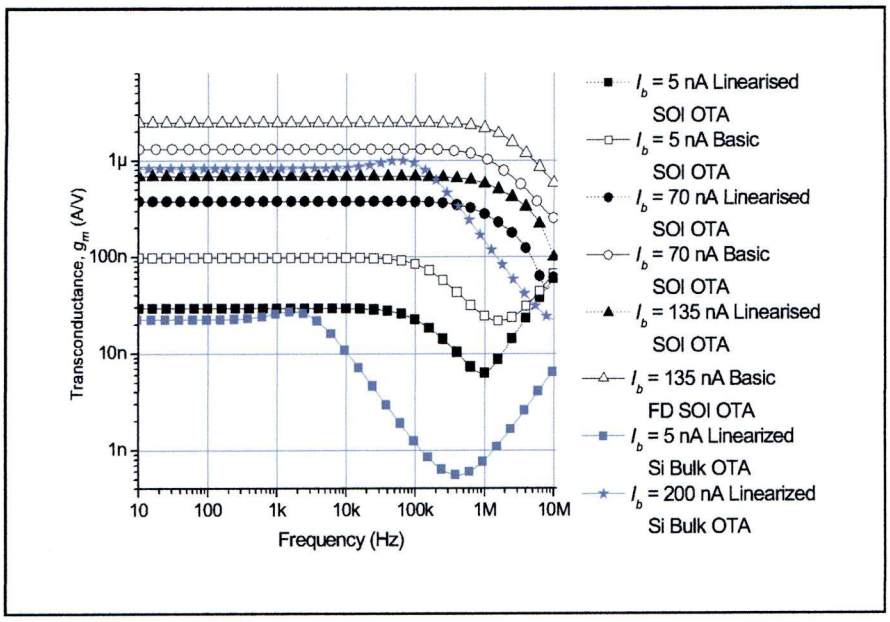

Figure 7.22, Transconductance bandwidth for basic and linearized OTA; using Si Bulk and FD SOI devices; at different bias currents.

The transconductance bandwidth was measured with a differential voltage amplitude of 10 mV to be within the 1% linear range for both circuits. Fig. 7.22 shows the FD SOI OTA has a larger bandwidth than the Si bulk OTA and the linearization reduces the transconductance bandwidth by 31% at a bias current of 5 nA and 4.1% at 200 nA. At 5 nA the bandwidth for the Si bulk OTA is 5.5 kHz, i.e. below the 20 kHz of audio. However for a cascade of filters with a range of cut off frequencies spaced on a log scale this range of bandwidths is ideal.

Table 7-4, Comparision of Si Bulk and FD SOI OTA in terms of; transconductance; 1% Linear differential voltage input; transconductance bandwidth for I_b = 1 nA to 100 nA.

		Si Bulk	FD SOI
Transconductance			
(I_b = 1 nA to 100 nA)	Basic	16 nA/V to 1.6 μA/V	45 nA/V to 1.8 μA/V
	Linearized	5 nA/V to 429 nA/V	8 nA/V to 555 nA/V
1% Linear differential voltage input range			
(I_b = 1 nA to 100 nA)	Basic	12 mV	10 mV
	Linearized	180 mV	120 mV
Transconductance bandwidth			
(I_b = 1 nA to 100 nA)	Basic	1.82 kHz to 130.9 kHz	70 kHz to 1.8 MHz
	Linearized	6.32 kHz to 501.5 kHz	53 kHz to 1.2 MHz

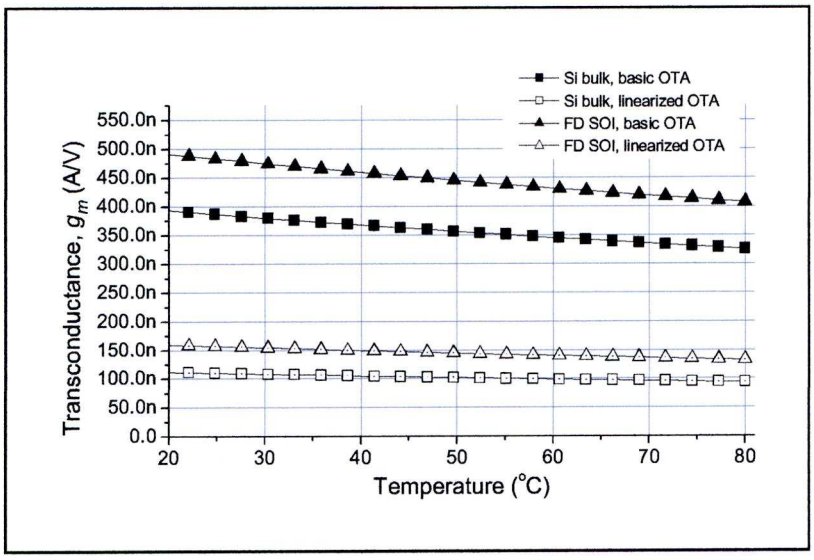

Figure 7.23, Dependence of transconductance, at V_{id} = 0 V, on circuit temperature for Si bulk and FD SOI MOSFETS in a basic and linearized OTA circuit, I_b = 25 nA.

Temperature simulations were performed for the circuits in the range 20^0C to 80^0C to investigate how the linearity would be affected by heating within the circuit. The EKV simulations for temperature (fig. 7.23) show the same linear effect for both FD SOI and Si bulk. The increased temperature reduces g_m, which would effectively increase the linearity. Therefore we can conclude there is no negative effect caused by temperature increases within the OTA.

7.9 Dynamic Range of G_m-C filter

The G_m-C filters use the OTA in closed loop configuration. A first-order low-pass follower integrator filter with transfer function $1/([C/g_m]+1)$ and cut-off frequency is given by g_m/C was simulated as described in section 7.8.1. The bias current of the OTA controls the cutoff frequency. The dynamic range is measured as the ratio of the intensities of the largest and smallest inputs the system can tolerate. The upper limit is limited by the non-linearity causing distortion at the output, while the lower limit is set by the intrinsic input-referred noise floor.

Total Harmonic Distortion: Typically the upper limit for the dynamic range of a first order G_m-C filter is strongly related to the linearity of the OTA. As the OTA is connected in a closed loop we also expect the distortion to be related to the g_m/I_D of the OTA. Thus when measuring the total harmonic distortion, (THD), at the cut-off frequency the reduced linearity of the OTA is balanced by an increase in the g_m/I_D when comparing FD SOI and Si bulk MOSFETS, as seen in table 7-5.

Table 7-5, Input rms voltage that produces 4% THD for a first order G_m-C filter.

	Si Bulk	FD SOI
Basic OTA	99.0 mV_{rms}	134.4 mV_{rms}
Linearized OTA	466.7 mV_{rms}	438.4 mV_{rms}

Noise: The input equivalent noise voltage was simulated with HSpice and the EKV models for FD SOI and Si bulk devices. The noise is calculated using the superposition of the flicker and thermal noise for each of the transistors transformed to the positive differential voltage input of the OTA as discussed in section 6.5. If the noise signal is referred back to the input, the reduction in 'n' improves the thermal noise in FD SOI over Si bulk and the increase in (g_m/I_D) lowers the flicker noise in FD SOI compared to Si bulk for similar transistors. Figure 7.24 shows the noise spectrum for FD SOI compared to Si bulk across the range of frequencies of interest for the first order G_m-C filter. Flicker noise can be viewed as producing noise at low frequencies; however at higher frequencies the

noise is almost exclusively thermal. Based on experimental measurements previous work on linearized Si bulk OTAs has indicated higher intrinsic noise than shown with our simulations [7.9]. For our investigation of dynamic range, the important result indicated by the simulation is the lower intrinsic noise achieved with FD SOI.

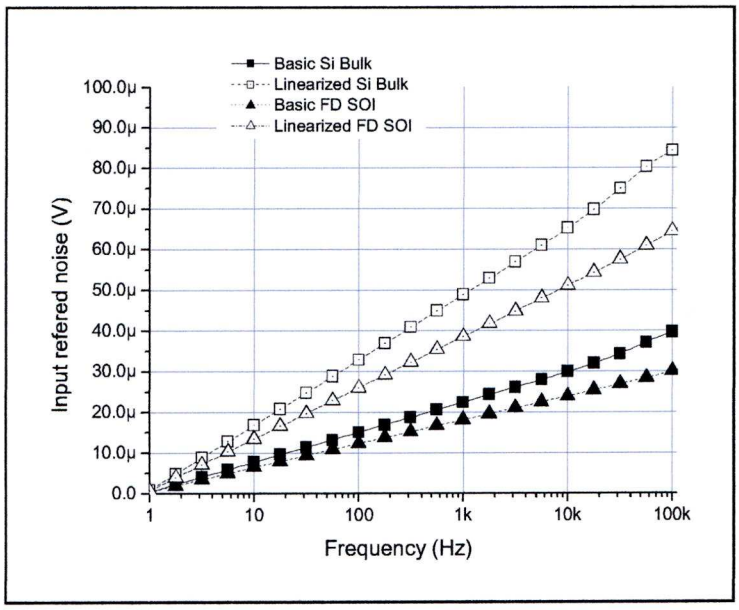

Figure 7.24, Equivalent input voltage noise for a first order G_m-C filter with a Si bulk and FD SOI MOSFETS basic and linearized OTA.

Table 7-6 shows the dynamic range for a first order G_m-C filter constructed with various OTAs. The dynamic range is quoted in logarithmic units of decibels (dB) and calculated from the intensity of the ratio of input signal that produces 4% THD and the rms input inferred noise. The dynamic range is improved by the linearization of the OTA and the improvement in noise performance increases the dynamic range of the FD SOI circuit compared to the Si bulk circuit.

Table 7-6, Dynamic Range for a first order G_m-C filter.

	Si Bulk	FD SOI
Basic OTA	68.0 dB	72.3 dB
Linearized OTA	76.0 dB	77.2 dB

7.10 Conclusion

This chapter discussed the design, modelling, simulation and development of an analogue filter with variable cut off frequency and adjustable gain for a hearing aid frequency feature extraction system. The building blocks of these filters are OTAs and capacitors. Hence these filters are called G_m-C filters. The implementation of the OTA is compared in Si bulk and FD SOI technologies. The OTA in FD SOI was implemented in silicon at the UCL laboratories in Belgium. This implementation was used to validate the modelling and simulation of the FD SOI technology. Following the validation of these models they were subsequently used to compare a FD SOI and a Si bulk simulation of a G_m-C filter.

The filters that were developed were 2^{nd} order G_m-C filters that uses three OTAs and two capacitors. The cut off frequency and gain is controllable via the adjustment of the transconductance of the OTAs. The transconductance is controlled by a bias current as discussed in 6.4.1.

Chapter 7 identifies the need to improve the dynamic range for low power circuits. This is needed to improve the peak to peak input voltage/supply voltage ratio to minimise power consumption. This requires improvement of the linear operating input voltage range of the OTAs. Hence an improved linearity OTA was developed using two techniques. These techniques were source degradation and bump linearization.

The effect of improving the linearity of the OTA on the dynamic range of G_m-C filters was studied for both FD SOI and Si bulk implementations. The traditional view would be that the increased transconductance of FD SOI devices compared to Si bulk devices, as discussed in 6.6.7, would significantly reduced the linear input voltage range of the OTA and therefore reduce the dynamic range of the filter. This thesis shows that there is a reduction in the linear input voltage range for FD SOI compared to Si bulk. This reduction is still shown when linearization techniques are used to improve the input range for both OTA implementations. However this thesis shows that the linear input range is only one of the factors

controlling dynamic range of the filter. The linear input range controls the upper limit of the dynamic range by providing a limit above which the total harmonic distortion significantly affects the output. The other lower limit of the dynamic range is the intrinsic noise level of the filter circuit. The thesis shows that the increased transconductance for FD SOI reduces the equivalent input intrinsic noise voltage of the filter compared to Si bulk. This is because the transconductance reduces the noise equivalent voltages when they are transformed by the circuit and referred back to the input. Therefore a study of the dynamic range for both the linearized FD SOI filter and the Si bulk filter shows that the increased transconductance does not, in fact, significantly affect the dynamic range of filters.

The G_m-C FD SOI analogue filter can therefore be developed as a cascade of filters with variable cut off frequencies and controllable gain. These can then form the frequency feature extraction system of a hybrid analogue/digital hearing aid system as developed through this thesis.

References

[7.1] Geiger, R. L., Sanchez-Sinencio, E., "Active Filter Design Using Operational Transconductance Amplifiers: A Tutorial", IEEE Circuits and Devices Magazine, vol. 1, pp. 20-32, Mar. 1985.

[7.2] Flandre, D., Gentinne, B., Eggermont, J. P., Jespers, P., "Design of Thin-film Fully-depleted SOI CMOS Analogue Circuits significantly outperforming Bulk Implementations", Proc. IEEE International SOI Conference, pp. 99-100, Oct. 1994.

[7.3] Flandre, D. et al., "Fully-depleted SOI CMOS technology for heterogeneous. micropower, high-temperature or RF microsystems", Solid-State Electronics, vol. 45, no. 4, pp. 541-549, 2001.

[7.4] Lyon, R. F., C. Mead, C., "An analog electronic cochlea", IEEE Trans. Acoust., Speech, Signal Processing, vol. 36, pp. 1119-1134, Jul. 1988.

[7.5] Watts, L., Kerns, D., Lyon, R., Mead, C., "Improved Implementation of the Silicon Cochlea", IEEE Journal of Solid-state Circuits, vol. 27, no. 5 , pp. 692-700, May 1992.

[7.6] Mead, C., "Analogue VLSI and Neural Systems", Addison Wesley Publishing Company, Inc., 1989.

[7.7] Chilakapati, U., et al., "A CMOS Transconductor with 80-dB SFDR up to 10Mhz", IEEE Journal of Solid-Stage Circuits, vol. 37, no. 3, Mar. 2002.

[7.8] Furth, P. M., Andreou, A. G., "Linearised differential transconductors in subthreshold CMOS", Electronic Letters, vol. 31, no. 7, pp. 545–547, Mar. 1995.

[7.9] Trung-Kien, N., Sang-Gug, L., "Low-Voltage, Low-Power CMOS Operation Transconductance Amplifier with Rail-to-Rail Differential Input Range", Proc. of IEEE International Symposium on Circuits and Systems (ISCAS' 06), Kos, Greece, pp. 4, May 2006.

[7.10] Sarpeshkar, R., Lyon, R., Mead, C., "A low-power wide-linear-range transconductance amplifier", Analog Integrated Circuits and Signal Processing, vol. 13, pp. 123-151, 1997.

[7.11] Delbruck, T., "Bump Circuits for Computing Similarity and Dissimilarity of Analogue Voltages", Proc. of IEEE International Joint Conference on Neural Networks, vol. 1, pp. 475-479, Seattle, Jul. 1991.

[7.12] Hasler, I., Buiu, O., Rue, B., Flandre, D., Hall, S., "Design of a low voltage – low power highly linear fully depleted SOI OTA for hearing aid applications", Proc. of EUROSOI 2006, pp. 27-28, Grenoble, Mar. 2006.

Chapter 8

8 ADAPTIVE FILTERS IMPLEMENTED AS A DSP CORE PROCESS

This chapter will move away from the frequency feature extraction hearing aid circuit and system discussed in chapters 6 & 7.

Chapter 8 will discuss a noise cancellation system. The requirement for this system within the hearing aid structure was identified and discussed in chapter 2 and can be seen within the context of an overall hearing aid in figure 2.4.

As a result of deficiencies in their auditory systems, subjects who are affected by hearing impairment find it difficult, or impossible, to perceive speech in the presence of noise [8.1]. Also subjects who already use a hearing instrument complain that in noise polluted environments, because of the way their hearing aids operate, they find it very difficult to perceive speech.

This chapter develops the idea of an intelligent hearing aid noise cancellation system, demonstrating how this can be implemented using adaptive filters. The chapter looks in detail at the use and functioning of adaptive filters in hearing aids, and important related issues such as mis-alignment and mis-adjustment. The proposed system is designed, modelled, and further tested with real life noise and speech to give a quantitative measure of its effectiveness.

A dual input adaptive noise cancellation system has been investigated because of its ability to adapt to different and varying noise environments, and its ability to discriminate between different and varying target signals. The system uses a simple but effective normalised LMS adaption algorithm. Issues of target mis-alignment and algorithm mis-adjustment/ convergence rate are discussed and suggestions for the development of the

system are presented. The addition of simple speech detection and the use of a varying convergence rate parameter have been investigated. The performance is analysed in terms of the overall energy improvement of the signal. A significant improvement, greater than 5 dB, of the log energy compared to the required target energy was obtained for signal-to-noise ratios in the range from –20 dB to 20 dB.

8.1 Adaptive Filters for Noise Cancellation

Filters can be used to remove noise in order to make speech more intelligible. Noise cancellation is however, not a simple filtering problem, since in general terms both speech and noise have time variant spectrums. Research in this area, to date, has shown the ability to significantly improve the quality of sound in terms of removing noise [8.2]; however this does not necessarily improve the intelligibility of the speech. Therefore the successful use of filtering to counteract noise interference will focus on improving speech intelligibility rather than just improving the signal to noise ratio.

8.1.1 Noise

There are three types of noise that can damage speech intelligibility;

***Random Noise*:** This is produced from road traffic, aeroplanes, industrial machinery, etc.

Secondary Interference voice or voices: This is sometimes call 'babble' and caused by speech from other people talking occurring at the same time as the primary speaker. The interference effect is also known as the 'cocktail party effect' where the sound of everybody talking at once makes it difficult to interpret an individual's speech. This situation can occur in many social situations such as bars or restaurants, etc.

Reverberation: This is produced by the speech sound being reflected off surrounding hard surfaces and the delayed reflected speech interfering with the direct speech from the speaker.

There are a number of issues that make noise cancellation for hearing aids a complex problem:

- The non-stationary nature of the noise.
- The wide range of noise and target speech frequency contents and amplitudes.
- The close and overlapping noise and target spectrum components causing frequency masking.

8.1.2 Adaptive Filters

An adaptive filter is simplistically a system filter with parameters and characteristics that vary over time. The variations in the adaptive filter parameters are controlled by an adaptive filter algorithm which describes how the behaviour of the filter should change over time. These changes will be in response to the different system requirements that are supplied to the adaptive algorithm in the form of training data.

8.1.3 Adaptive Filter Structure

As shown in figure 8.1, the filter is effectively continuously redesigned to attempt to create an optimal filter based on the knowledge of the required signal *d(n)*. Therefore as the filter is continuously redesigned, the filter structure can have a relatively simple structure. This structure can be a simple FIR filter with coefficients that can be redesigned by the filter's algorithm.

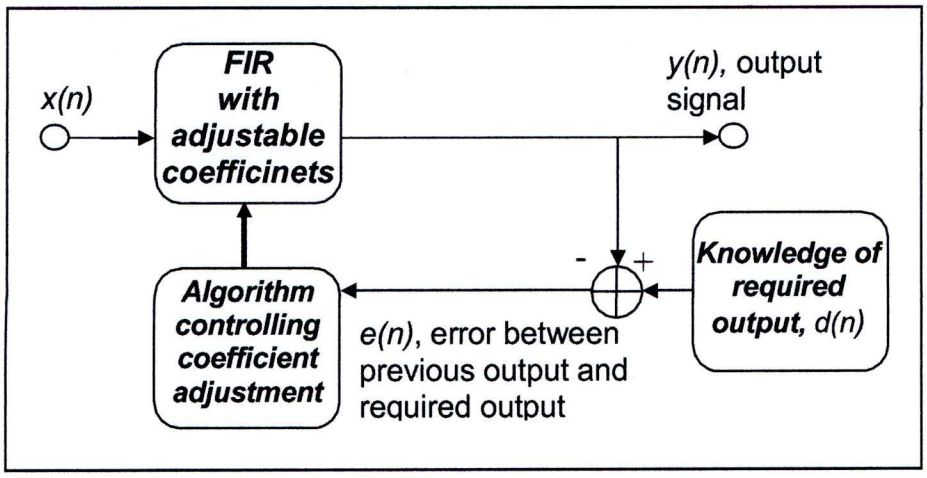

Figure 8.1, Adaptive Filter Structure.

A FIR filter is straight forward to implement and can be easily adapted with an algorithm.

8.2 Adaptive Filter Algorithms

There are a number of different algorithms that have been developed for adaptive filters.

This section of the chapter will show the work that has been carried out investigating the algorithm that is used to adjust the adaptive filter. This uses work carried out by Widrow and Stearns in the field of adaptive signal processing [8.3] and Haykin whose work provides a more in-depth review of adaptive filters [8.4].

8.2.1 The LMS (Least Mean Squares) Algorithm

The least-mean-squares (LMS) algorithm is highly popular and is widely used in a variety of applications. It was originally proposed by Widrow and Hoff in 1960 [8.5] and it is the algorithm that is used as the basic tool for the study and development of adaptive filters. The LMS algorithm is simple to implement in computer models, yet is powerful enough to provide a good evaluation of the benefits of an adaptive filter within an application.

8.2.2 Recursive Least Squares (RLS) Algorithm

The RLS algorithm uses a recursive equation for updating the deterministic correlation function of the filter coefficients, such that given the least-squares estimate of the coefficients vector of the filter for the last input data sample, the updated coefficients can be calculated for the next input data sample.

The RLS algorithm has a faster convergence rate and produces a more accurate steady state result compared to the LMS algorithm; however these advantages have to be weighted against the additional computation that is required for this adaptive algorithm.

8.2.3 Evolutionary Filters Using Genetic Algorithms

This type of filter uses Darwinian theories to develop the optimum system solution based on the biological processes of reproduction from two independent parents and random mutations to produce a new population. This new population is then reduced to the best suited population for further reproduction and mutation. The reduction is controlled by what are called 'environmental' factors based on the principle of only the most suitable or strongest in the population being able to survive and reproduce; in practical terms the environment being the requirements of the system. From the continued evolution of the solution, the optimum system solution is presented and arrived at in a way that would be impossible to emulate as a simple algorithm.

Ade, M. et al. [8.6] have developed a practical implementation for this type of filter. In numerical simulations the filters have been shown to have a higher convergence rate and smaller steady-state value of the MSE compared to LMS adaptive filters. However, similarly to the RLS algorithm, the genetic algorithms require much greater computational resources to adapt the filter.

8.2.4 Adaptive Algorithm Conclusion

It was concluded that compared to other algorithms the LMS algorithm's simplicity and its ability to still provide good results in terms of adaption of the filter, make it a good choice to develop for a low power hearing aid application.

8.3 The LMS Algorithm

The LMS algorithm uses the 'mean-squared error', *MSE* or $\xi(i)$, taken from the 'error signal', *e(i)*, or the difference between the training signal/desired response, *d(i)*, and the actual filter output, *y(i)*. The algorithm implements a 'Stochastic Gradient Approach' for minimising the 'mean-squared error' to produce an approximation of the optimum Weiner solution for the filter coefficients of a FIR filter. The variation of the *MSE* for a varying filter coefficient vector, w_i, forms a multidimensional paraboloid with a defined minimum point. Using the derivative of the MSE, an iterative approach can be used to modify the filter coefficients towards the optimum. A parameter must also be introduced to control the rate at which the descent towards the minimum *MSE* occurs. This parameter is called the step-size parameter, μ_{ssp}.

Appendix C shows the LMS algorithm function is,

$$w_{i+1} = w_i + \mu_{ssp} E\{e(i)x^*(i)\} \tag{8.1}$$

Or in simple terms,

$$\begin{pmatrix}\text{Updated values}\\ \text{of filter}\\ \text{coeffcicents}\end{pmatrix} = \begin{pmatrix}\text{Previous values}\\ \text{of filter}\\ \text{coeffcicents}\end{pmatrix} + \begin{pmatrix}\text{Step - Size}\\ \text{parameter, } \mu_{ssp}\\ \text{(Learning Rate)}\end{pmatrix}\begin{pmatrix}\text{Error}\\ \text{Signal}\end{pmatrix}\begin{pmatrix}\text{Input data}\\ \text{vector}\end{pmatrix}$$

8.3.1 Step-size Parameter

The step-size parameter or convergence factor, μ_{ssp}, controls the convergence rate of the 'Stochastic Gradient Approach' towards the minimum. It also controls the stability of the approach and hence the stability of the overall adaptive algorithm.

The subsequent section describes the work carried out to investigate how the step-size parameter affects the convergence rate or learning curve of the algorithm.

8.3.2 *Stability of LMS Algorithm*

Hayes, M [8.7] describes how the adaptive filter converges if,

$$0 < \mu_{ssp} < 2/\lambda_{max}$$

where, λ_{max} is the maximum eigenvalue of the autocorrelation matrix R_{xx},

$$R_{xx} = E\{xx^H\} \tag{8.2}$$

and where x^H denotes the complex conjugate transpose of x the input data vector.

8.3.3 *Analysis of the LMS algorithm*

The LMS adaptive filter was simulated in Matlab with a simple adaptive noise canceller set-up as shown in figure 8.2. The noise added to the desired signal was white noise with a variance of 0.1. In order to easily analyse the effect of each parameter of the LMS algorithm a simple time invariant unitary input signal was initially used with the simulation.

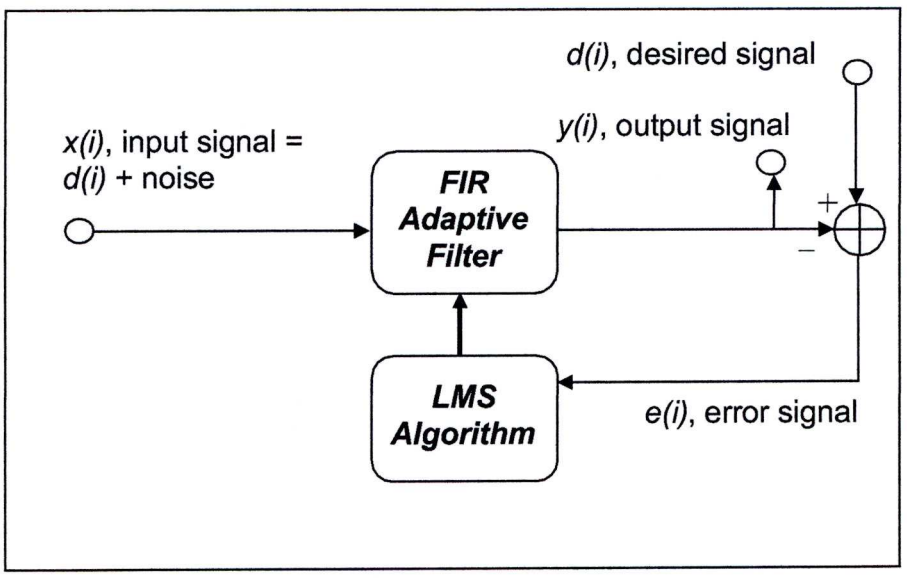

Figure 8.2, Set-up of a simple Adaptive Noise Canceller.

8.3.4 Time invariant unitary input

Figure 8.3 shows how the adaptive filter trains itself to eliminate the noise and approximate the unitary input signal over a number of consecutive samples of the input. With each sample a new iteration of the LMS algorithm is calculated and an updated approximation of the unitary input signal from the adaptive filter is produced, $y(i)$.

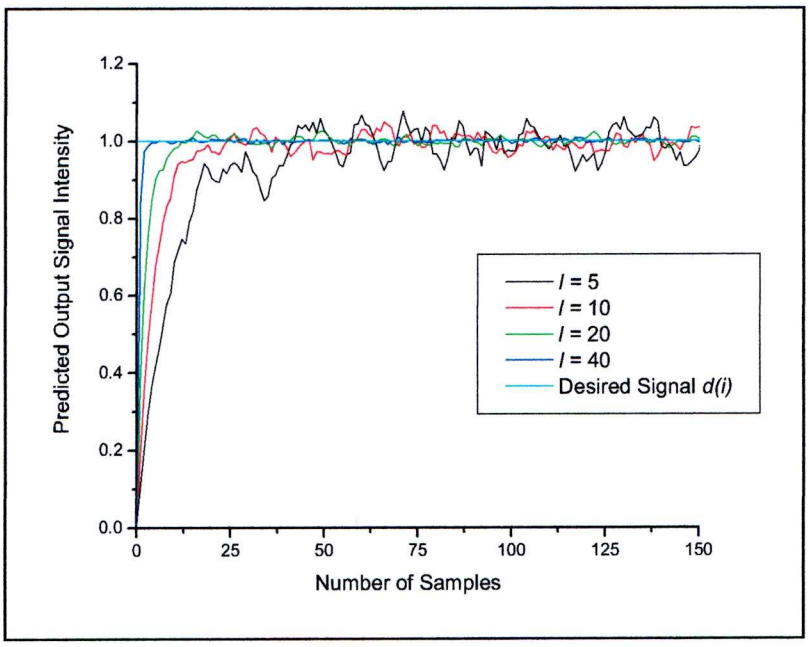

Figure 8.3, Predicted output signal intensity of LMS Noise Cancellation Simulation for Unitary Input with varying Filter Lengths, l, with $\mu_{ssp} = 0.01$.

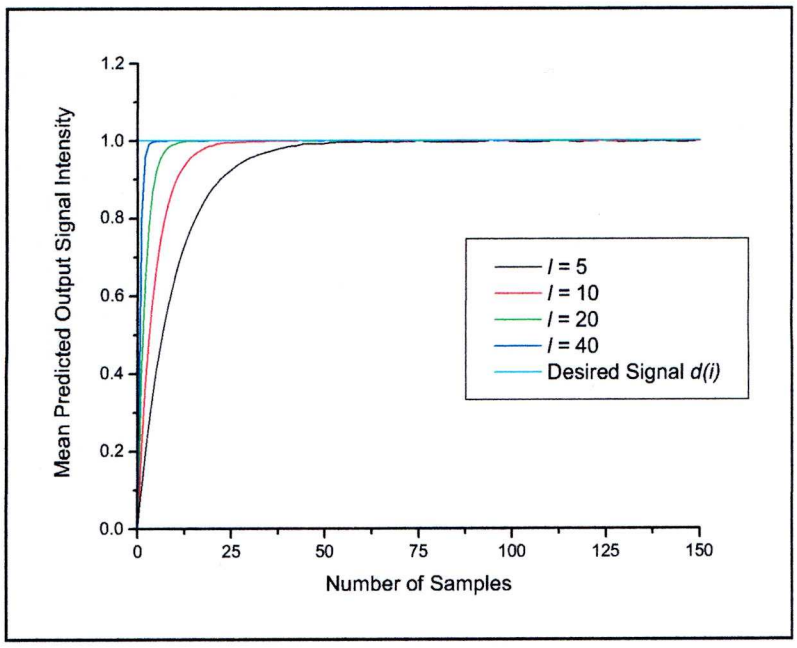

Figure 8.4, Mean predicted output signal intensity of 1000 independent trials of LMS Noise Cancellation Simulation for Unitary Input with varying Filter Lengths, l, with $\mu_{ssp} = 0.01$.

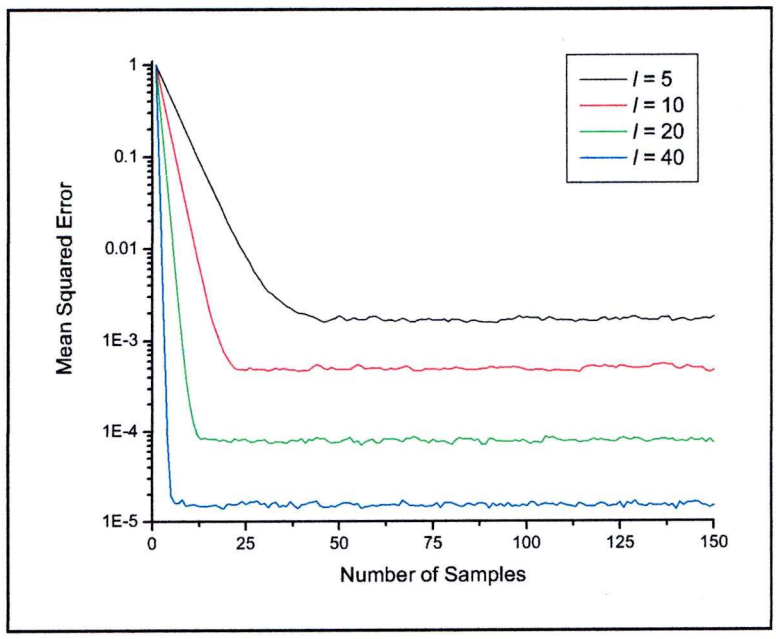

Figure 8.5, MSE for 1000 independent trials of LMS Noise Cancellation Simulation for Unitary Input with varying Filter Lengths, l, with $\mu_{ssp} = 0.01$.

8.3.5 Number of filter coefficients (Filter Length, l)

Figure 8.3 shows the output of the filter *y(i)* for 150 input samples which produces 150 iterations of the adaptive filter coefficients, where l is the

number of coefficients of the FIR filter and the step-size parameter is fixed at 0.01.

Figure 8.3 shows the output for a single trial with a single set of input samples *x(i)* for each of the filter lengths *l*. From figure 8.3 it can be observed that the transient behaviour of the output follows a noisy exponential curve.

By taking the ensemble average of the output over 1000 independent trials the effect of the exponential curve noise can be cancelled out as shown in figure 8.4.

From figure 8.4 it can be observed that the algorithm has an exponential convergence between the output, *y(i)* and the desired signal *d(i)*. As the filter length is increased the rate of convergence is increased giving the filter a faster response; at the same time a reduction in the error in the output after the initial convergence has occurred is recorded.

The final point stated above was further analysed and results presented in figure 8.5. This is a plot of the MSE that is produced for 1000 independent trials against the number of input samples or iterations of the LMS algorithm and is repeated for various values of filter length. The MSE is calculated from the error, *e(i)*, which is the difference between the output, *y(i)*, and the desired input, *d(i)*.

We can observe from figure 8.5 the convergence rate or 'learning curve' of the algorithm as the MSE approaches its settled value. An increase in the filter length will result in a decrease of the convergence period, as well as a decrease of the steady state value reached by the MSE.

Now it can be stated that as the filter length increases the convergence/learning period is reduced and the accuracy of the output, after the learning period, is increased. Hence the filter is faster acting and more accurate. This however has to be balanced with the fact that the greater the filter length, the more hardware resources are required and the

longer the filter will need to operate. To further investigate the analysis of the effects of the filter length, figure 8.6 shows the settled MSE, after the convergence period, i.e. after $100 - l$ iterations of the LMS algorithm, against filter length.

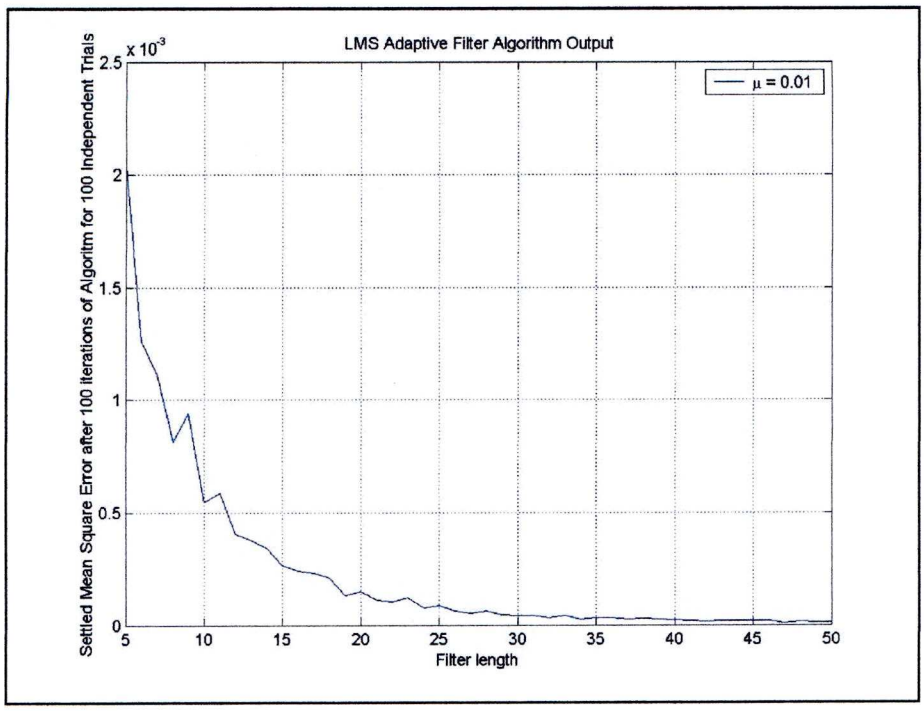

Figure 8.6, Settled MSE against Filter Length, l, for a LMS Noise Cancellation with Unitary Input, $\mu = 0.01$.

The errors in the curve shown in figure 8.6 are due to the noise in the settled MSE after $100 - l$ samples and 100 independent trials. This noise can be observed also in figure 8.5 and could be reduced by taking the assembled average for more than 100 independent trials.

An initial increase in the filter length ($l < 10$), has a large effect on the reduction in the MSE, but as the filter length becomes large, $l > 25$, there is little gain in terms of the MSE.

Figure 8.7 is a plot of the time constant of the learning curve, which can be observed in figure 8.5, for the MSE averaged over 1000 independent trials against the filter length.

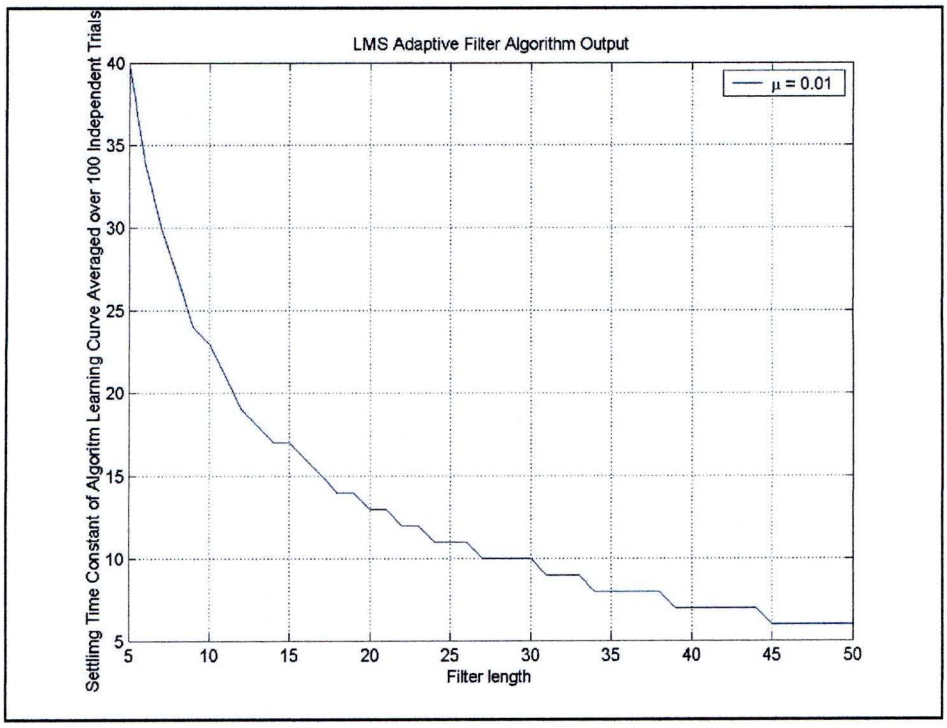

Figure 8.7, MSE Learning Curve against Filter Length, l, for a LMS Noise Cancellation with Unitary Input Input, $\mu = 0.01$.

From figure 8.7 it can be seen that the learning curve or convergence time of the adaptive filter reduces as the filter length increases; however this has a decreasingly significant effect in reducing the convergence time as the filter length increases. Therefore to balance the computational overheads for a longer filter with the performance requirements, a filter length of approximately 10 is required.

8.3.6 Step-size Parameter, μ_{ssp}

Figure 8.8 shows the output of the filter *y(i)* for 500 input sample which produces 500 – l iterations of the adaptive filter coefficients, where l is the number of coefficients of the filter and is fixed at 10.

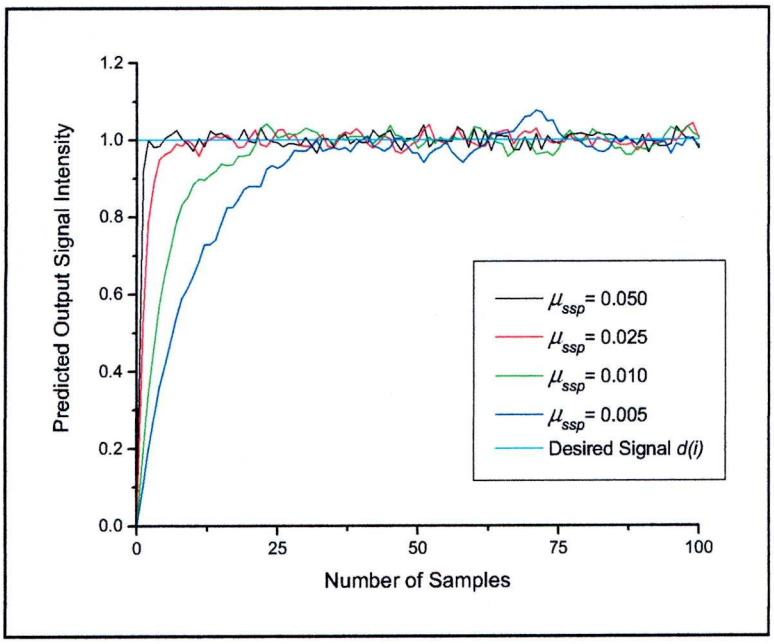

Figure 8.8, Predicted output signal intensity of LMS Noise Cancellation Simulation for Unitary Input with varying μ_{ssp} and filter length $l = 10$.

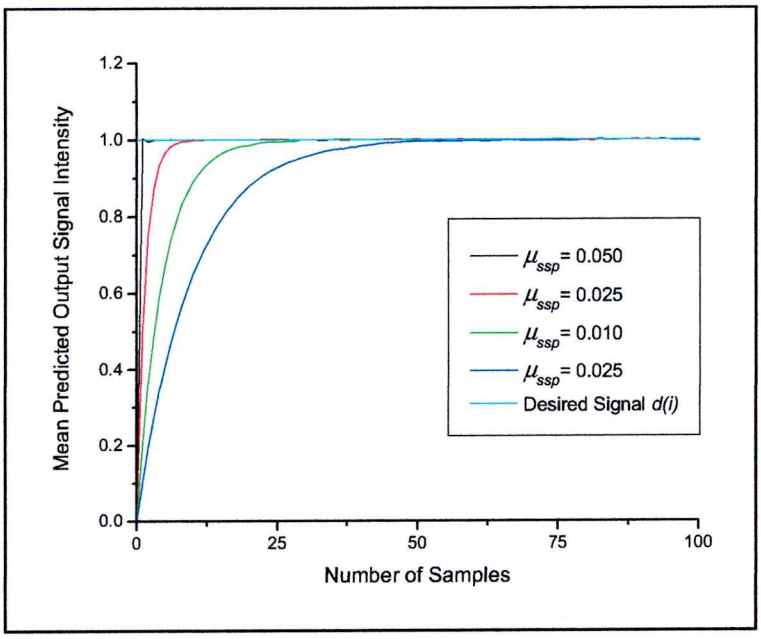

Figure 8.9, Mean predicted output signal intensity of 1000 independent trials of LMS Noise Cancellation Simulation for Unitary Input with varying μ_{ssp} and filter length $l = 10$.

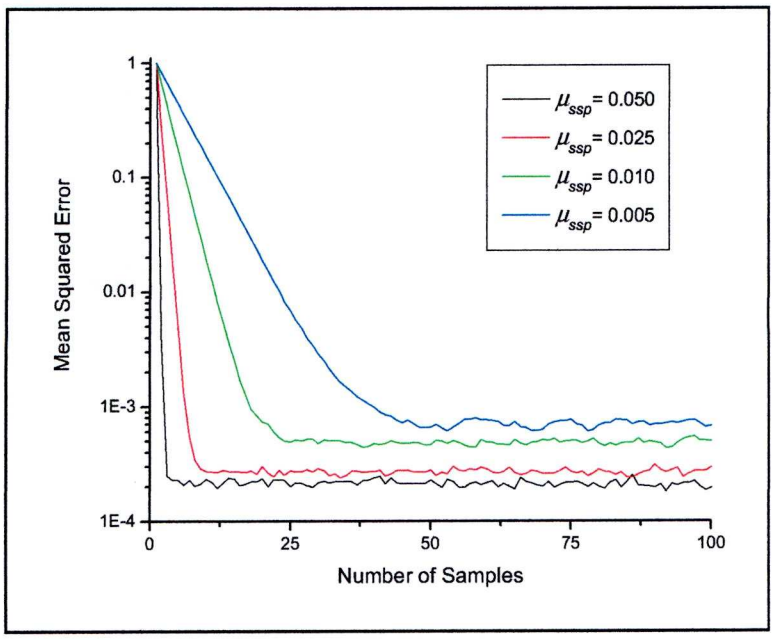

Figure 8.10, MSE for 1000 independent trials of LMS Noise Cancellation Simulation for Unitary Input with varying μ_{ssp} and filter length $l = 10$.

Figure 8.8 shows the output for a single trial with a single set of input samples *x(i)* for each tested value of μ_{ssp}. From figure 8.8 it can be observed that the transient behaviour of the output follows a noisy exponential curve.

By taking the ensemble average of the output over 1000 independent trials the effect of the exponential curve noise can be cancelled out as shown in figure 8.9.

From figure 8.9 it can be observed that the algorithm has a convergence between the output, *y(i)* and the desired signal *d(i)*. As μ_{ssp} is increased the rate of convergence is increased giving the filter a faster response: According to the theory, given in Haykin [8.4] for this input data, the maximum value of μ_{ssp} for a stable algorithm should be approximately 0.18. Thus there is no instability that can be observed when the maximum $\mu_{ssp} = 0.05$.

Figure 8.10 is a plot of the MSE that is produced for 1000 independent trials against the number of input samples or iterations of the LMS algorithm for various values of μ_{ssp}. From this figure 8.10 it can be

observed that for μ_{ssp} increasing from 0.005 to 0.05, the steady state value of the MSE decreases and the convergence rate increases, resulting in a faster and more accurate adaptive filter.

8.3.7 Normalised LMS Algorithm

The normalised LMS algorithm is a modified version of the previously described LMS algorithm. A significant problem with the LMS algorithm is that in order to calculate the limit of stability for the step-size parameter, full knowledge of the complete filter input data must be known. However, for most practical applications, this is not known in advance. Hence the normalised LMS uses a modified step-size parameter. This new parameter is based on the assembled squared average of all input samples up until the present sample,

$$\hat{E}\left\{|x(i)|^2\right\}$$

Thus the Normalised LMS has the function for updating the coefficients [8.8],

$$w_{i+1} = w_i + SF\frac{x^*(i)}{\|x(i)\|^2}e(i) \tag{8.3}$$

where SF is a stability factor and for stability in all cases of input data,

$$0 < SF < 2$$

8.4 Systems using Adaptive Filters for Noise Reduction

Following the study of the adaptive filter this thesis can now move on to look at practical systems for noise cancellation in hearing aids using adaptive filters.

A typical application for an adaptive system is the separation of stochastic and deterministic contributions in a signal. Therefore it would appear they are well suited to the requirements of hearing aids to remove time variant noise if the speech is viewed as a deterministic signal.

A variety of adaptive single microphone noise cancellation techniques have shown reasonable success; however the most successful systems require frequency domain operation to estimate the localised noise spectrum which adds significantly to the power consumption of a system. Multiple-element sensor systems have produced good noise and target discrimination and also have the advantage that they can, to some extent, discriminate signals spatially.

8.4.1 Short Term Invariant Noise

Figure 8.11a shows a simple adaptive noise reduction strategy in which a speech detector allows the adaptive filter coefficients to be updated during localised periods of non-speech [8.2]. It is then assumed that the noise spectrum is a short-term invariant so that during periods of speech the filter can be fixed to remove the previously determined noise spectrum from the speech.

8.4.2 Adaptive Noise Cancellation

An adaptive noise canceller, (ANC) is shown in figure 8.11b. This uses a two-stage adaptive 'beamformer' scheme as proposed by D. Van Compernolle [8.8]. This scheme has shown significant improvement (on average more than 5 dB) in speech reception threshold in background noise [8.9]. ANC's such as this require the use of multiple directional microphones or a microphone array. It is assumed that all microphones receive non-directional noise and one microphone receives significantly more of the directional speech than the other microphones. Hence the difference in the inputs is used to adapt and cancel out the noise from the speech. The two-microphone 'beamforming' configuration also has the advantage of spatially filtering noise and providing a directional selection for the primary speech input in addition to simply filtering the speech from non-speech noise; however the strategy does require the use of multiple microphones.

8.4.3 Feedback Cancellation

Figure 8.11c shows an adaptive filter used in a feedback cancellation strategy. The adaptation of the filter is delayed so that the filter is adapted to remove the previous output of the filter. Therefore when the feedback from the previous output is received at the input microphone as reverberation noise the filter is adapted to remove it. This type of strategy, using a low-power adaptive filter, was described by Kim, et al, [8.10].

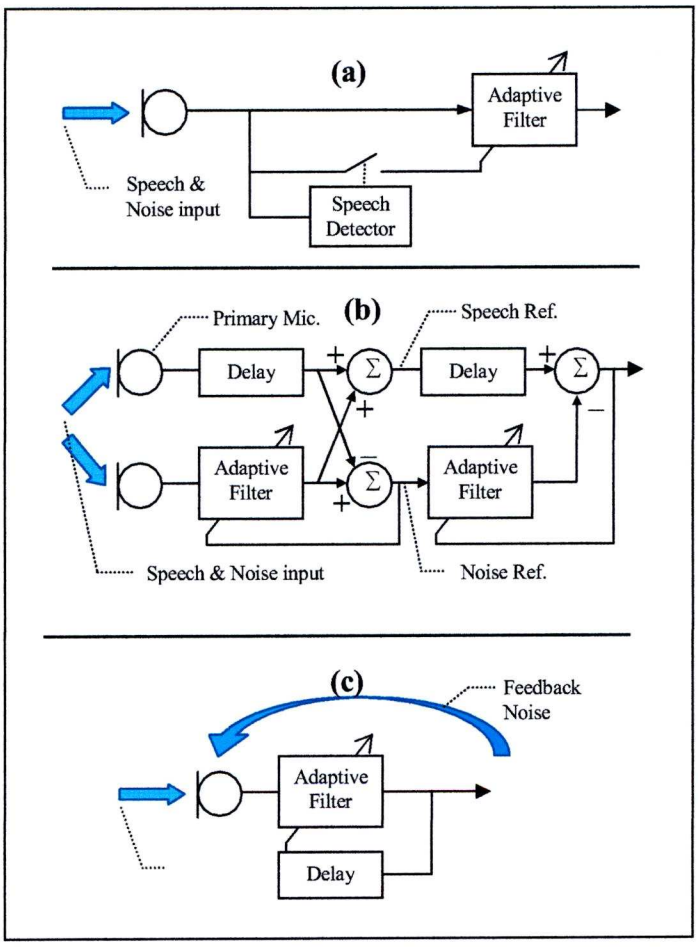

Figure 8.11, Adaptive Filter Noise Reduction Structures, a) A simple short term invariant noise canceller using speech detection and an adaptive filter, b) A dual microphone noise canceller with two adaptive filters, c) feedback noise cancellation with a delayed adaptive filter.

8.5 Dual Microphone Adaptive Noise Cancellation

In this thesis the performance of an adaptive two-microphone noise-reduction system is evaluated and developed. Similar systems have already shown the possibility of providing significant speech enhancement for hearing aid applications [8.10].

The system studied here is based on an adaptive beamformer algorithm as described by Griffith and Jim [8.11]. A basic two-microphone version of the system for a straight in-front speech target is shown in figure 8.12. The sum and difference of the microphones is used to provide primary and reference inputs for a classical adaptive noise canceller using a finite impulse response (FIR) digital adaptive filter as described by Widrow and Stearns [8.3]. The difference, or reference input, ideally provides an input signal containing noise only due to the subtraction and cancellation of the target speech signal, whilst the sum or primary input contains target and noise. The reference signal passes through a FIR filter whose weights are adjusted to minimise the error in the system output. This minimization is achieved by the filtering of the reference signal to correlate with the noise signal in the primary path and then taking the difference of both. As the target and noise are uncorrelated the minimization has no effect on the target signal, which results in an output signal with minimum noise and no distortion to the speech. The addition of a delay in the primary path can improve performance by allowing both past and future filtered reference samples to contribute to the output.

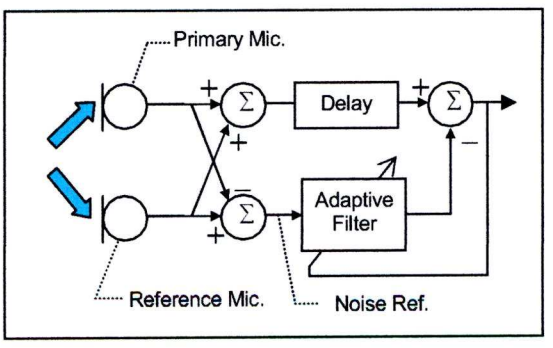

Figure 8.12, Simple Two-microphone adaptive noise cancellation beamformer schematic.

The algorithm used to adapt the FIR filter weights is the normalised LMS algorithm, as discussed in section 8.3.7. The selection of the stability factor, *SF*, is critical to the performance of the ANC. The adaptive algorithm must adapt and converge fast enough to remove the relatively low modulation noise, whilst not adapting too fast so that the algorithm adapts to remove the fast modulating target speech.

8.5.1 Method

The evaluation of the system was carried out using synthetically mixed speech and noise. For the result shown here, the speech is a male voicing the sentence "the police helped the driver", recorded at 22.5 kHz in a reverberating real-room. The noise was recorded inside a car travelling at ~ 65 mph along a major road. The noise and speech were mixed at varying signal-to-noise ratios, SNR, and the reference microphone input was supplied with a signal with a SNR ~ 2 dB less than the primary microphone input signal.

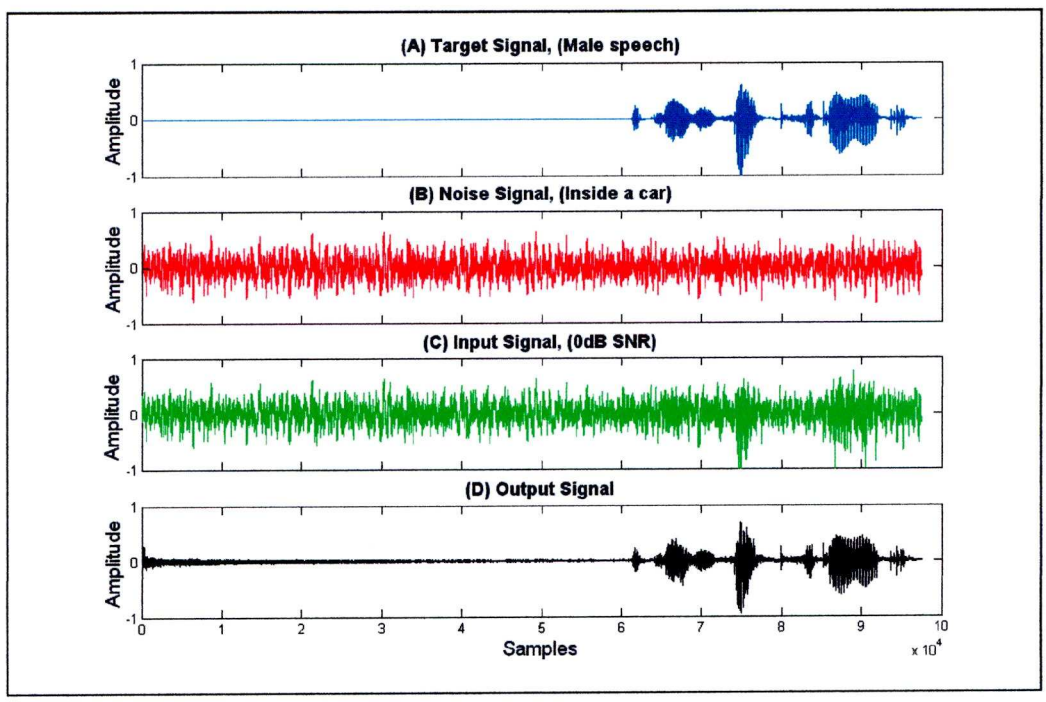

Figure 8.13, Simulation results for Simple Two-microphone Adaptive Noise Canceller (from top to bottom: a) Target Speech, (Male speaker), b) Noise Signal,(Inside a car), c) Input Signal, (0 dB SNR), d) Output Signal).

Figure 8.13a shows the target speech input signal with a period of silence at the start which is used to evaluate the convergence rate of the system. Figure 8.13b shows the noise input signal for the same period and figures 8.13c & 8.13d show the system input signal and the resulting output signal.

The performance of the system was measured in terms of the system's ability to remove noise energy, whilst preserving target signal energy overall, for 10 ms sample windows of the signal. The principle of superposition was used to observe the system's effect on the target signal energy and the jammer signal energy separately, and hence the overall energy improvement of the system. This provided a more valuable and consistent performance analysis than simply investigating the change in the SNR since degradation of the target signal can cause significant errors in the SNR results.

Figure 8.14 shows the energy improvement for the system at – 20 to 20 dB SNR. In this idealised case of complete cancellation of the noise signal in the filtering path, there is total preservation of the target signal. Since there is no change in the target energy, the overall energy reduction is the same as the reduction in the jammer noise signal. At high noise levels there is a significant reduction in the energy of the jammer up to – 46 dB at -20 dB SNR. As the jammer in the input signal reduces, the energy reduction of the jammer reduces until at 10 dB there is an increase in signal energy. This loss of performance with reduced jammer energy is caused by high frequency noise resulting from the adaptation of the filter. As the input jammer/noise energy reduces, the filter added noise becomes more prevalent until at approximately 8 dB SNR it becomes greater than the jammer energy the filter removes.

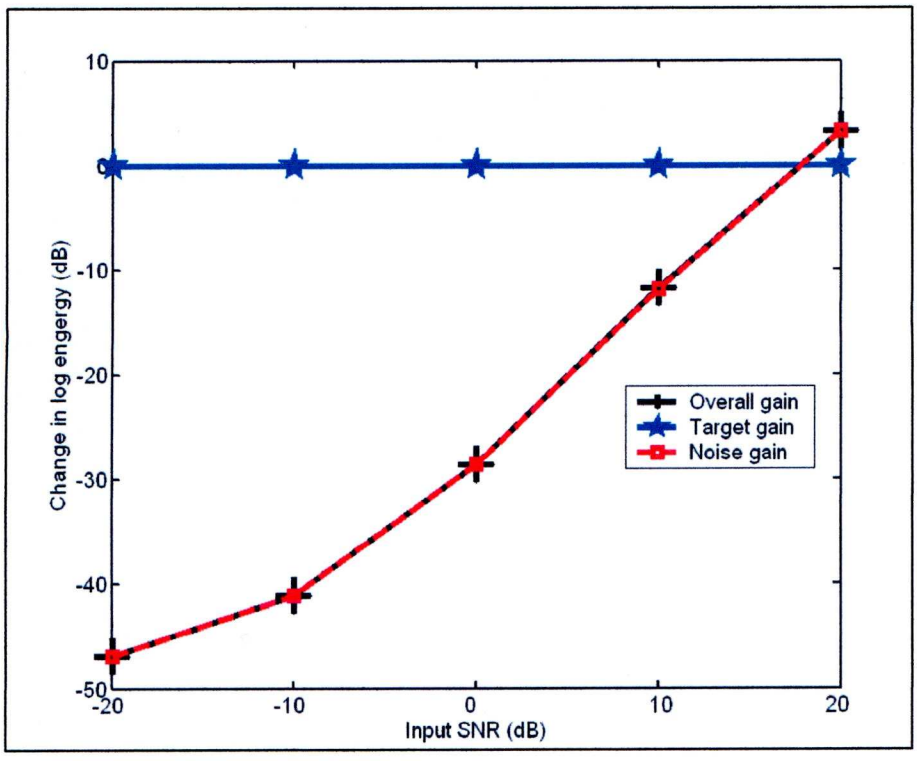

Figure 8.14, Energy improvement for Simple Two-microphone Adaptive Noise Canceller for in-car noise with speech.

There are some important issues that must be discussed regarding the simple two microphone adaptive noise canceller.

8.5.2 Misalignment

The arrangement of the system evaluated above is an idealised case because there is complete cancellation of the target signal in the filtering path. The output is degraded by leakage of the target signal into the filtering path. This occurs if there is misalignment between the target signals picked up at one input compared to the other. Figure 8.15 shows the energy improvement results for the system with a misalignment of 10% between the target amplitude at one input compared to the other.

Also compared to the case without misalignment, the jammer energy reduction still shows similar characteristics for low SNR. However, filter noise caused by adjustment of the filter to the leak target signal adds up to 5 dB to the target signal which degrades the overall performance of the system. At higher SNR levels, as the leaking target energy becomes more

dominant than the noise in the filter path, the filter adapts to cancel the target signal and as such removes target energy whilst increasing jammer/noise energy.

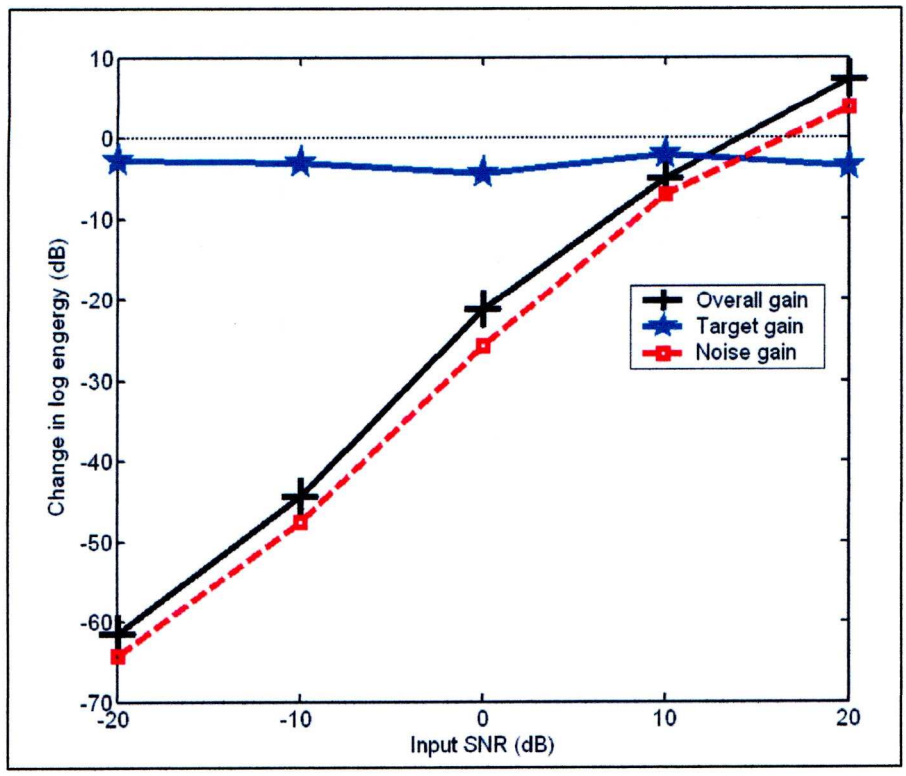

Figure 8.15, Energy improvement for Simple Two-microphone Adaptive Noise Canceller for in-car noise with misaligned speech.

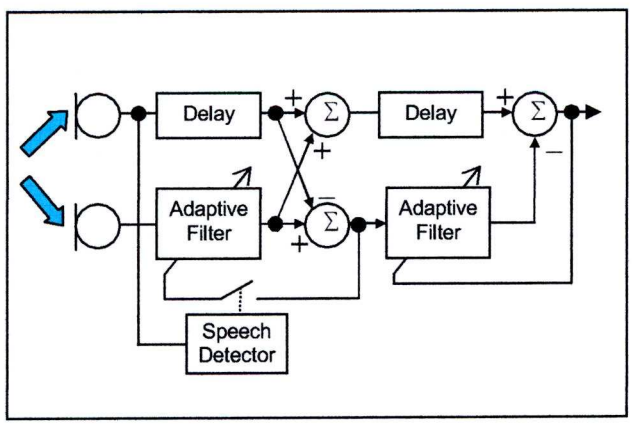

Figure 8.16, Two-microphone adaptive noise cancellation beamformer schematic with speech detection to eliminate misalignment.

Figure 8.16 shows the development of the system used to overcome the problem of misalignment. A simple speech detector is used to determine

when there is a predominance of target speech in the signal. During these periods a second adaptive filter is allowed to adapt to compensate for the misalignment. The energy improvement results for this system are shown in figure 8.17. For all SNR levels the overall performance is greatly improved and the effect of misalignment is no longer apparent. The overall performance gain is caused by the extra delay before the sum and difference are calculated. This delay and the initial adaptive filter further decorrelate the target and jammer signals used for the second adaptive filter. As such the performance of the second adaptive filter is improved. The delay is also responsible for the slight loss of target energy for all SNR levels.

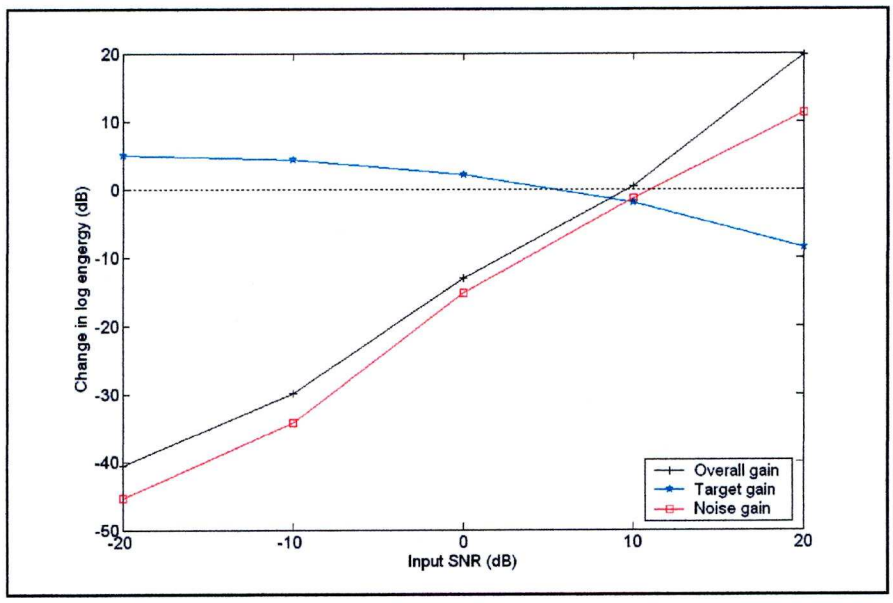

Figure 8.17, Energy improvement for Two-microphone Adaptive Noise Canceller with Speech Detection for in-car noise with misaligned speech.

8.5.3 Misadjustment

As the convergence parameter is increased the filter will adapt at a faster rate so that each time it adapts it will be closer to its optimum performance, thus cancelling more of the noise. However, when there is a target signal present the increased convergence rate will also result in an increase in the amount of misadjustment of the filter so that it adapts to cancel both the target and the noise. The noise cancellation performance

of the filter, is therefore, limited by the level at which misadjustment degrades the target signal.

To overcome the issue of misadjustment the system has been developed to have a variable convergence rate parameter. With an optimum convergence rate parameter and no target signal the filter adapts quickly to produce a highly effective cancellation filter for the noise. However, when the target signal appears, because it is highly uncorrelated with the noise, it causes rapid, sudden changes in the filter weight vector causing misadjustment. By monitoring the rate of change of the adaptive filter weight vector, when misadjustment occurs, the convergence rate can be reduced to zero and the filter weight vector, that was optimised before the misadjustment occurred, used to filter the noise whilst preserving the target. Figures 8.18 and 8.19 show the energy improvement results for a variable convergence rate system. Compared to figure 8.17 there is a significant improvement in the overall performance due to the equally significant reduction of the jammer, greater than – 35 dB for all SNR levels.

The Variable Convergence Rate ANC shows excellent performance characteristics. However, there are issues still to be resolved regarding performance with multi jammer speech type inputs and consideration of the balance between performance and computation complexity and the issues of fast speech onset to avoid loss of initial speech sound.

8.6 Conclusion

This chapter discussed the development of an adaptive noise cancellation system for a hearing aid as identified in the hearing aid structure developed in figure 2.4.

The chapter discussed the use of adaptive filters for adaptive noise cancellation. Adaptive filter algorithms have also been studied and discussed. The LMS and normalised LMS are identified as the most

appropriate algorithm to use because of its simplicity yet good adaption performance.

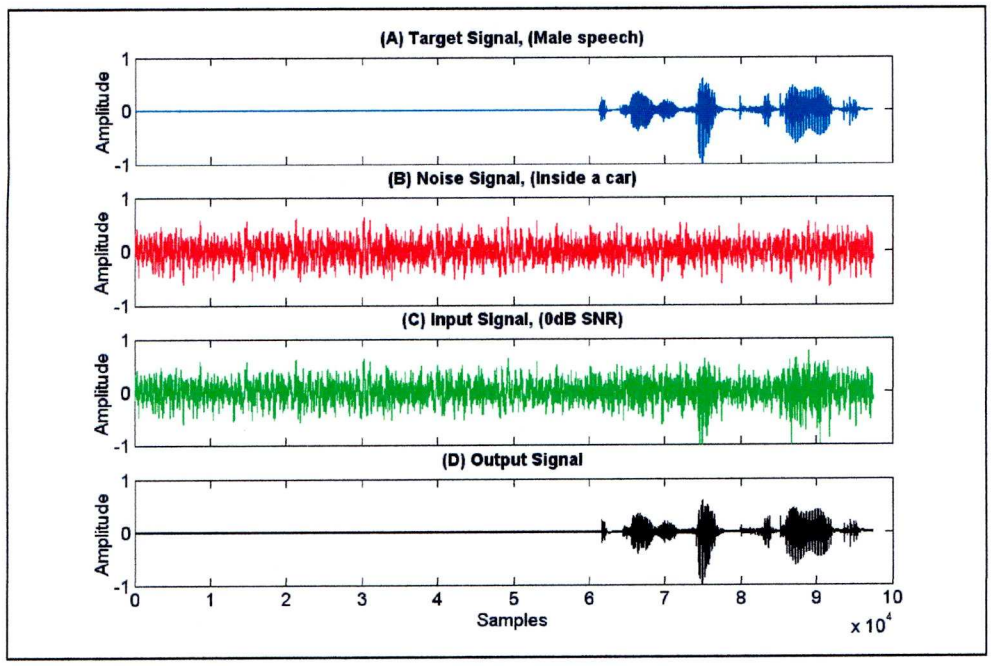

Figure 8.18, Simulation results for Two-microphone Adaptive Noise Canceller with Speech Detection for in-car noise with misaligned speech, (from top to bottom: a) Target Speech, (Male speaker), b) Noise Signal,(Inside a car), c) Input Signal, (0 dB SNR), d) Output Signal).

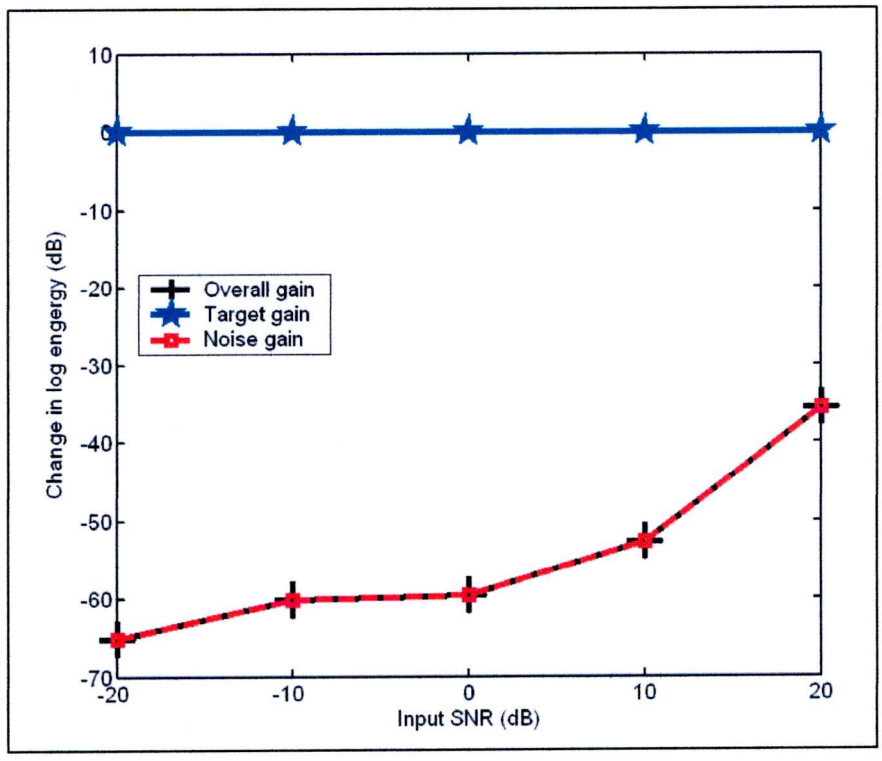

Figure 8.19, Energy improvement for Two-microphone Adaptive Noise Canceller with Variable Convergence Parameter for in-car noise with speech.

The trade–off between computation requirements, i.e. filter length, and performance is identified. The best filter length is shown to be approximately 10.

The chapter then discussed the development of a dual microphone adaptive noise cancellation system. The system is developed to overcome issues of misalignment and misadjustment. Misalignment is overcome with the addition of a second adaptive filter. Misadjustment is overcome with the addition of a speech detection algorithm and a variable convergence parameter.

The developed adaptive noise cancellation system is shown to have a good performance across a wide range of SNRs. The performance was analysed in terms of the overall energy improvement of the signal. A significant improvement, greater than 5dB, of the log energy compared to the required target energy was obtained for signal-to-noise ratios in the range from – 20 dB to 20 dB.

The adaptive noise cancellation system developed in this chapter can be implemented as a system within a complete hybrid analogue/digital hearing aid system.

References

[8.1] Kollmeier, B., http://political.hear-it.org/multimedia/kollmeyer.pdf "Futural Trends in Hearing Aid Technology", University of Oldenburg, 2003.

[8.2] Levitt, H., "Noise reduction in hearing aids: An overview", Journal of Rehabilitation Research and Development, vol. 38, no. 1, pp. 111-121, Feb. 2001.

[8.3] Widrow, B., Stearns, S. D., "Adaptive Signal Processing", Prentice Hall, Englewood Cliffs, NJ, 1985.

[8.4] Haykin, S., "Adaptive Filter Theory", Prentice Hall, Englewood Cliffs, 3rd ed., 1996.

[8.5] Widrow, B., Hoff, M. E., "Adaptive Switching Circuits", IRE-WESCON Conv. Rec., vol. 4, pp. 96-104, Aug. 1960.

[8.6] Abe, M., Kawamata, M., "Hardware implementation of evolutionary digital filters", Proc. of the 2003 International Symposium on Circuits and Systems, vol. 4, pp. 393-396, May 2003.

[8.7] Hayes, M. H., "Statistical Digital signal Processing and Modelling", John Wiley & Sons, Inc., NY, 1996.

[8.8] Van Compernolle, D., "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings", Proc. of International Conference on Acoustics, Speech and Signal Processing, ICASSP 90, vol. 2, pp. 833-836, Abuquerque, Apr. 1990.

[8.9] Berghe, J. V., Wouster, J., "An adaptive noise canceller for hearing aids using two nearby microphones", J. of the Acoust. Soc. of America, vol. 103, no. 6, pp. 3621-3626, Jun. 1998.

[8.10] Kim, C. H. I., Soeleman, H., Roy, K., "Ultra-low power DLMS adaptive filter for hearing aid applications", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 11, no. 6, pp. 1058-1067, Dec. 2003.

[8.11] Griffiths, L. J., Jim, C. W., "An alternative approach to linearly constrained adaptive beamforming", IEEE Transactions on Antennas and Propagation, vol. 30, no. 1, pp. 27-34, Jan. 1982.

Chapter 9

9 SUMMARY AND CONCLUSIONS

This thesis has presented research into the electronic circuits and system that would be required for a new generation of hearing aids. From the start of this project, on which this thesis is based, the area of research was very open ended, because this was the first work to be carried out on this research area within the Electronic Department at the University of Liverpool. Therefore one of the aims of the project was to identify what opportunities were available to contribute new work and knowledge to the field of hearing aid circuits and systems.

The natural human auditory system is used as the basis to form a design for an electronic hearing aid system to compensate for deficiencies in the natural human auditory system. The thesis shows the natural human auditory system to be highly efficient and because any hearing aid will have to integrate effectively with the natural human auditory system it appears logical to use it as a starting point for a hearing aid design. The aim of a hearing aid has to be to improve communication so, in parallel with the initial investigation into the human auditory system and hearing impairment, there was an analysis of human speech and noise from a signal processing perspective.

The initial research enabled the human auditory system to be broken down into a model of blocks each with a particular function. This model was then used as the basis for the different systems required for a hearing aid.

To create a specification for hearing aid circuits and systems this thesis presents a study of current hearing aid technology circuits and systems. From this research the major issues for improving hearing aids were identified as the following:

- Improving low power consumption to enable longer battery life.

- Use of a very low voltage to limit battery size.
- Being highly miniaturised to help with aesthetics,
- Improving functionality to provide the best signal processing possible.
- Increasing flexibility to allow good fitting to a user's need in different environments.
- Have a high signal to noise ratio, to enable the user to perceive speech easily without noise interference.
- Have low distortion of speech to enable the user to communicate and understand speech effectively.

This thesis concentrates on the development of the frequency feature extraction system within the hearing aid model. This system was chosen to be developed because it was highlighted as the central component of most overall systems that would be required for a large range of hearing aid types and to overcome a large range of hearing impairments.

The challenge discussed in this thesis then moved on to evaluating the best technology for implementing a frequency feature extraction system and circuits within a hearing aid.

It has been concluded that hybrid mixed analogue/digital signal processing would be the best solution for hearing aid circuits and system. This conclusion was drawn because a hybrid approach uses the power efficiency advantages of both analogue signal processing and digital signal processing whilst limiting the disadvantages of both. It was also highlighted that a hybrid signal processing approach has many parallels with the highly efficient natural human auditory system's signal processing approach. As previously mentioned the hearing circuits and system designs presented at the start of this thesis were highly influenced by the natural human auditory system so continuing the inspiration this thesis

shows there are strong engineering reasons for using the hybrid processing approach for hearing aid circuits and systems.

The project also focused on previous work into modelling the human auditory system as an electronic system and in particular the functionality of the cochlea. This led to the development of the hybrid Lyon & Mead [9.1] model using a cascade of low pass filters which implement a frequency feature extraction system within a hearing aid. The Lyon & Mead model was chosen because of its excellent power consumption characteristics using sub-threshold VLSI analogue CMOS [9.2].

The use of a hybrid mixed signal processing approach leads to the development of silicon on insulator (SOI) technology to implement hearing aid circuits because of its performance with a mixed signal system and the possibilities for silicon micro-machined (MEM) microphones to provide a complete system-on-chip.

An opportunity to contribute work within the field of SOI technology for analogue circuits was identified, because although previous work had shown the benefits of SOI technology for analogue circuits [9.3] there is little published work in the area.

The project focussed on the development of G_m-C filters for a frequency feature extraction system implemented in fully-depleted (FD) SOI technology. The fundamental building block of the filter was therefore the operational transconductance amplifier (OTA). The OTA was designed for FD SOI technology, which involved the modelling FD SOI using suitable HSpice models. An OTA was designed, simulated, laid out, verified, fabricated and tested with work carried out at UCL, Microelectronics Laboratory, Louvain-la-Neuve, Belgium. This work demonstrated the use of the technology for the circuit and verified the modelling of FD SOI.

The need to improve the linearity of OTAs within G_m-C filter circuits for hearing aids was identified. There has been previously published work on

improving the linear input range for Si bulk technology, but this is not directly applicable to FD SOI technology. The traditional view was that the increase in transconductance for FD SOI devices compared to Si Bulk devices would decrease the dynamic range of an FD SOI OTA. Therefore new work on increasing the linearity of a FD SOI OTA is presented and is fully compared to a similar Si bulk OTA. This work was presented at EUROSOI 2006 [8.4]. The work was used to show how, when operating in a closed loop configuration within a G_m-C filter circuit, the FD SOI OTA has no overall disadvantages over the Si bulk OTA in terms of dynamic range. This is shown to be because the effect of increased transconductance for FD SOI devices occurs at the same time as a reduction in the intrinsic equivalent input noise compared to a Si bulk OTA.

The final work presented in this thesis discusses the development of powerful but energy efficient digital algorithms to achieve flexible and adaptive signal processing which will have major performance advantages for the hearing aid user.

As a separate part of the hearing aid model, discussed earlier, a two microphone adaptive noise cancellation (ANC) system based on previous work, [9.5], was developed. The ANC system was shown to provide effective noise suppression as a constituent part of an intelligent hearing aid. The use of a variable convergence rate parameter produced the best performance for a wide range of SNRs; however, there are issues still to be resolved regarding performance with multi jammer speech type inputs and establishing the right balance between performance and computation complexity. This work was presented by the author at the CAS2004 conference [9.6].

This thesis enables and justifies the continuation of work into designing further circuits and systems based on the hearing aid model presented herein. It also justifies the continued use of SOI technology for a hybrid mixed analogue/digital signal processing approach for these further circuits and systems. This work into hearing aid circuits and systems can be

combined with other research into micromachined microphone on SOI technology in particular. The ultimate goal of this work can then be to achieve a powerfully performing, small scale, power efficient, flexible and adaptive hearing aid device as a signal system-on-chip solution.

References

[9.1] Lyon, R. F., C. Mead, C., "An analog electronic cochlea", IEEE Trans. Acoust., Speech, Signal Processing, vol. 36, pp. 1119-1134, Jul. 1988.

[9.2] Vittoz, E. A., "Very low Power Circuit Design: Fundamentals and Limits", Proc. of ISCAS'93, pp. 1439-1442, Chicago, May 1993.

[9.3] Flandre, D., Vanhoenacker, D., "Thin-film fully-depleted SOI CMOS technology, devices and circuits for LVLP analogue/digital/microwave applications", Proc. of Semiconductor Conf, CAS '98, vol. 1, pp. 115-124, 1998.

[9.4] Hasler, I., Buiu, O., Rue, B., Flandre, D., Hall, S., "Design of a low voltage – low power highly linear fully depleted SOI OTA for hearing aid applications", Proc. of EUROSOI 2006, pp. 27-28, Grenoble, Mar. 2006.

[9.5] Berghe, J. V., Wouster, J., "An adaptive noise canceller for hearing aids using two nearby microphones", J. of the Acoust. Soc. of America, vol. 103, no. 6, pp. 3621-3626, Jun. 1998.

[9.6] Hasler, I., Buiu, O., "Design of a noise cancellation system for intelligent hearing aids", Proc. Int. CAS Conf. 2004, vol. 2, pp. 479-482, 2004.

Appendix A

For a MOSFET device operating in weak inversion the gate voltage induces a small positive surface potential, Ψ_s, all along the channel. This biases the np junction at the source and electrons are injected into the channel. However these few electrons are enough in number to give a detectable current flow but their associated charge is insufficient in magnitude to influence the electrostatics vertically through the device. In the channel the charge is insufficient to create a lateral field and electrons flow from source to drain by a process of diffusion. Therefore we can say,

$$I_D = qD_e \frac{vd_e}{c} \tag{A1}$$

where, c, is the distance along the channel, vd_e, is the volume density of electrons, D_e, is the diffusion coefficient for electrons and, q, is the charge on an electron.

If at the source (c = 0) and at the drain (c = L), the volume density of electrons injected at the source is $vd_e(0)$.

Therefore from (A1),

$$I_D = qD_e \frac{vd_e(0)}{L} \tag{A2}$$

It is then necessary to relate $vd_e(0)$ to the surface potential by considering the energy bands vertically into the substrate at the source, c = 0.

Well known p-n junction theory enables us to write,

$$vd_e(0) = vd_{eo} \exp\left(\frac{\Psi_s}{U_T}\right) \tag{A3}$$

where vd_{eo} is the equilibrium number of electrons in the substrate, dependant on the substrate doping and intrinsic carrier concentration, and U_T is the thermal voltage.

Substiting (A3) into (A2) gives the surface potential, drain current relationship,

$$I_D \propto \exp\left(\frac{\psi_s}{U_T}\right) \tag{A4}$$

Hence if we can relate surface potential to V_{GS} we can relate I_D to V_{GS} in weak inversion. This is achieved by examining the electrostatics of the MOS capacitors formed by the gate oxide and substrate, as shown in figure A1 exists. Considering the electrostatic continuity of displacement across the oxide, silicon interface we can write,

$$\varepsilon_o \varepsilon_{ox} \xi_{ox} = \varepsilon_o \varepsilon_s \xi_s \tag{A5}$$

where ξ_{ox} and ξ_s are the electric fields in oxide and silicon respectively and ε_o, ε_{ox}, ε_s, are the permittivities of free space, the oxide layer and silicon layer respectively.

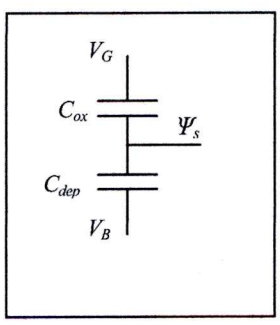

Figure A1, The body effect coefficient Si bulk MOSFET.

The oxide field is given by,

$$\xi_{ox} = \frac{V_{GS} - \psi_s}{t_{ox}} \tag{A6}$$

where t_{ox} is the gate oxide thickness.

For the silicon, from Poisson's equation,

$$\xi_s = \frac{d\psi_s}{dh} \tag{A7}$$

where h is the distance into the silicon substrate.

Therefore,

$$\psi_s = \frac{\xi_s h}{2} \tag{A8}$$

We also require a relationship between field and charge. This is provided by Gauss' Law,

$$\xi_s = \frac{qN_a h}{\varepsilon_o \varepsilon_s} \tag{A9}$$

where N_a is the substrate doping.

Substituting for h in equation (A8) and (A9) gives,

$$\xi_s = \sqrt{\frac{qN_a \psi_s}{\varepsilon_o \varepsilon_s}} \tag{A10}$$

Substituting (A6) and (A10) into (A5) gives,

$$V_{GS}(\psi_s) = t_{ox} \frac{\varepsilon_s}{\varepsilon_{ox}} \sqrt{\frac{2qN_a \psi_s}{\varepsilon_o \varepsilon_s}} + \psi_s \tag{A11}$$

Equation (A11) is approximately plotted in figure A2.

For figure A2 the linear region of interest between **A** and **B** can be approximated by,

$$V_{GS} = n\psi_s + C \tag{A12}$$

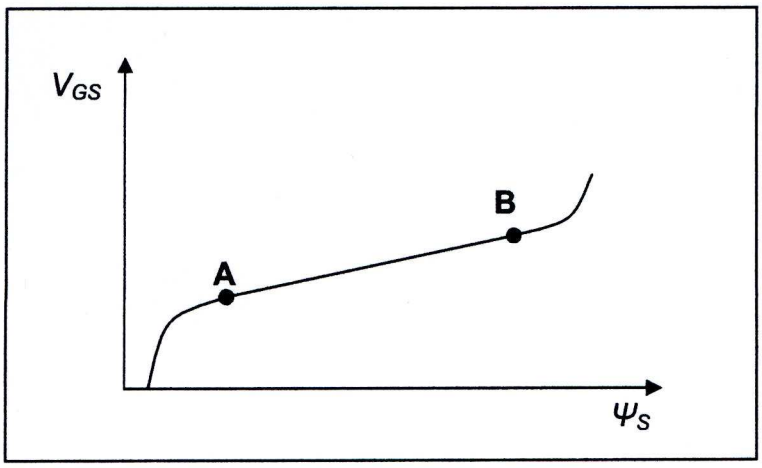

Figure A2, Plot of equation A11.

Substiting (A12) into (A4) we get the equation stated earlier,

$$I_D \propto \exp\left(\frac{V_{GS}}{nU_T}\right) \tag{A13}$$

The body coefficient, n, can be found by differentiating and equating (A11) and (A12),

$$n = \frac{dV_G}{d\psi_s} = \frac{t_{ox}}{\varepsilon_o \varepsilon_{ox}} \sqrt{\frac{q\varepsilon_o \varepsilon_s}{2\psi_s}} + 1 \tag{A14}$$

This can be written in a simpler form as,

$$n = 1 + \frac{C_{dep}}{C_{ox}} \tag{A15}$$

Appendix B

From figure 7.6, where the C_3 value can be ignored due to its negligible size.

Summing the current at X

$$0 = g_{m1}\frac{V_{id}}{2} + V_X sC_1 + V_X\left(g_{ds1} + g_{ds3} + g_{m3}\right)$$

Since $g_{m3} >> g_{ds1} + g_{ds3}$

$$0 = g_{m1}\frac{V_{id}}{2} + V_X sC_1 + V_X g_{m3}$$

$$\therefore V_X = -\frac{g_{m1}V_{id}/2}{\left(g_{m3} + sC_1\right)}$$

Also

$$0 = -g_{m2}\frac{V_{id}}{2} + V_X g_{m4} + V_{out} sC_2 + V_{out}\left(g_{ds2} + g_{ds4}\right)$$

$$0 = -g_{m2}\left(V_{id}/2\right) + -\frac{g_{m4}g_{m1}\left(V_{id}/2\right)}{\left(g_{m3} + sC_1\right)} + V_{out}\left(sC_2 + g_{ds2} + g_{ds4}\right)$$

$$V_{out} = \frac{\left(\frac{V_{id}}{2}\right)\left(g_{m2} + \frac{g_{m4}g_{m1}}{\left(g_{m3} + sC_1\right)}\right)}{\left(sC_2 + g_{ds2} + g_{ds4}\right)}$$

$$\frac{V_{out}}{\left(V_{id}/2\right)} = \frac{\left(g_{m2} + \frac{g_{m4}g_{m1}}{\left(g_{m3} + sC_1\right)}\right)}{\left(sC_2 + g_{ds2} + g_{ds4}\right)}$$

Assuming $g_{m1} = g_{m2}$ & $g_{m3} = g_{m4}$

$$A_v = \frac{g_{m1}\left(1 + \frac{1}{\left(1 + sC_1/g_{m3}\right)}\right)}{\left(sC_2 + g_{ds2} + g_{ds4}\right)}$$

This gives the transfer function

$$A_v = \frac{g_{m1}}{g_{ds2} + g_{ds4}} \frac{\left(1 + \frac{sC_1}{(2g_{m3})}\right)}{\left(1 + \frac{sC_1}{g_{m3}}\right)\left(1 + \frac{sC_2}{g_{ds2} + g_{ds4}}\right)}$$

$$A_v = A_{vo} \frac{\left(1 - \frac{s}{z}\right)}{\left(1 - \frac{s}{p_2}\right)\left(1 - \frac{s}{p_1}\right)}$$

where

$$p_1 = -\frac{g_{ds2} + g_{ds4}}{C_2} \cong -\frac{g_{ds2} + g_{ds4}}{C_L}$$

$$p_2 = -\frac{g_{m3}}{C_1}$$

$$z = --\frac{2g_{m3}}{C_1}$$

$$p_1 < p_2 < z$$

Appendix C

The *MSE* is given by the expected value, $E\{.\}$ of the squared difference between the desired signal *d(i)* and the actual output signal *y(i)*.

Thus,

$$MSE = \xi(i) = E\left\{|e(i)|^2\right\} = E\left\{|d(i) - y(i)|^2\right\}$$

The updated filter coefficients w_{i+1} are given by the present coefficient less the gradient of the *MSE* times the step-size parameter μ_{ssp}, where ∇ is the gradient operator,

$$w_{i+1} = w_i - \mu_{ssp} \nabla \xi(i)$$

Next,

$$\nabla \xi(i) = E\left\{e(i) \nabla e^*(i)\right\}$$

It follow that the gradient of the error, $\nabla e(i)$, is given by the inverse of *x(i)*, and x^* denotes the transposed matrix.

$$\nabla e^*(i) = -x^*(i)$$
$$\nabla \xi(i) = -E\left\{e(i) x^*(i)\right\}$$

This then gives the LMS algorithm function as,

$$w_{i+1} = w_i + \mu_{ssp} E\left\{e(i) x^*(i)\right\}$$