DEPARTMENT OF ELECTRONIC AND ELECTRICAL ENGINEERING
LOUGHBOROUGH UNIVERSITY

ARTEFACT REDUCTION IN PHOTOPLETHYSMOGRAPHY

BY

MATTHEW JAMES HAYES, MEng

A doctoral thesis
submitted in partial fulfilment of the requirement for the award of
doctor of philosophy of Loughborough University

November 1998

Supervisor: Professor Peter R. Smith, Ph.D.
Department of Electronic and Electrical Engineering

For Catherine,

and in memory of my Grandfather, Albert Leslie Kelsey.

ABSTRACT

The use of optical techniques in biomedical monitoring and diagnosis is becoming increasingly widespread, primarily because of the non-invasive nature of optically derived measurements. Physiological analysis is usually achieved by characterisation of the spectral or temporal properties of the interaction between light and the anatomy. Although some optical measurements require complex instrumentation and protocols, recent technological advances have resulted in robust and compact equipment that is now used routinely in a multitude of clinical contexts. Unfortunately, these measurements are inherently sensitive to corruption from dynamic physical conditions or external sources of light, inducing signal artefact. Artefact is the primary restriction in the applicability of many optical measurements, especially for ambulatory monitoring and tele-medicine.

The most widely used optical measurement is photoplethysmography, a technique that registers dynamic changes in blood volume throughout the peripheral vasculature and can be used to screen for a number of venous disorders, as well as monitoring the cardio-vascular pulse wave. Although photoplethysmographic devices are now incorporated into many patient-monitoring systems, the prevalent application is a measurement known as pulse oximetry, which utilises spectral analysis of the peripheral blood to estimate the arterial haemoglobin oxygen saturation. Pulse oximetry is well established as an early warning for hypoxia and is now mandatory under anaesthesia in many countries. The problem of artefact is prominent in these continuous monitoring techniques, where it is often impossible to control the physical conditions during use.

This thesis investigates the possibility of reducing artefact corruption of photoplethysmographic signals in real time, using an electronic processing methodology that is based upon inversion of a physical artefact model.

Initially, the physical origins of the photoplethysmographic signals are explored by specific interpretations of a general heuristic model describing the coupling of light from distinct anatomical components. These interpretations are used to highlight the advantages in utilising non-linear transduction techniques, which are shown to re-normalise the signal dynamics with optical information. A general description of signal artefact is then developed within the framework of this heuristic model, aided by an experimental investigation. An inversion for artefact reduction is then obtained, with the resultant solution being simplified by the non-linearity. The consequences of this non-linear artefact reduction technique for subsequent signal analysis are discussed, culminating in a modified formulation for pulse oximetry that not only has reduced sensitivity to artefact but also possesses increased generality. The design and construction of a practical electronic system is then used to explore both the implementation issues and the scope of this technique. The performance of artefact reduction obtained is then quantified under realistic experimental conditions, demonstrating that this methodology is successful in removing or reducing a large proportion of artefact encountered in clinically relevant situations.

It is concluded that non-linear artefact reduction can be applied to any photoplethysmographic technology, reducing interpretation inaccuracies that would otherwise be induced by signal artefact. It is also speculated that this technology could enable the use of photoplethysmographic systems in applications that are currently precluded by the inherent severity of artefact.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Professor Peter R. Smith, for his guidance, support and friendship throughout the duration of my research. I would also like to thank all the other members of the Optical Engineering Group for their companionship and numerous technical discussions.

I must also thank the technicians and other staff from Loughborough University for their advice and assistance in construction of the electronic circuitry.

I also acknowledge the financial support of both the Engineering and Physical Sciences Research Council and the Defence Research Agency.

Finally, I wish to express appreciation of all my family and my friends, for their support, inspiration and encouragement, without whom this personal achievement would have little value.

CONTENTS

ABSTRACT ..... **i**

ACKNOWLEDGEMENTS ..... **iii**

CONTENTS ..... **iv**

CHAPTER 1
INTRODUCTION **1**

1.1 THESIS OVERVIEW 2

1.2 PHOTOPLETHYSMOGRAPHY 4

1.2.1 Principle of operation 4

1.2.1a Transmission and reflection modes 4

1.2.1b PPG signal components 5

1.2.1c What does PPG measure? 6

1.2.2 Major applications 7

1.2.2a Venous haemodynamic testing 8

1.2.2b Pulse oximetry 9

1.3 PULSE OXIMETRY 10

1.3.1 Principle of operation 11

1.3.2 The Beer-Lambert law 11

1.3.2a Logarithmic solution 13

1.3.2b Linear solution 14

1.3.3 Practical considerations 15

1.3.3a Logarithmic solution 15

1.3.3b Linear solution 16

1.3.4 Sensing periods 17

1.3.5 Calibration 18

1.4 LIMITATIONS OF THE PPG MODEL 19

1.4.1 Artefact 20

1.4.1a Motion artefact 21

1.4.1b Ambient artefact 22
1.4.1c Other artefacts 25
1.4.2 The clinical problem of PPG artefact 25
1.4.2a Corruption of pulse oximetry output 26
1.4.2b Applicability of the PPG model 26
1.5 EXTANT TECHNIQUES FOR ARTEFACT REDUCTION IN PULSE OXIMETRY 27
1.5.1 Correlation cancellation or signal processing 28
1.5.2 Feature based recognition of corrupt pulses 29
1.5.3 Cancellation via control of an additional light source 30
1.5.4 Inversion of an empirical artefact model 31
1.5.5 Conclusions 32
1.6 SUMMARY 34

CHAPTER 2
MODELLING 36

2.1 A HEURISTIC PPG MODEL 37
2.1.1 Definition of the model 37
2.1.2 Applicability 39
2.2 INTERPRETATION OF THE MODEL 39
2.2.1 Choice of physical model 40
2.2.2 Beer-Lambert interpretation 41
2.2.2a Separation of signal components 41
2.2.2b Interpretation 42
2.2.3 Diffusion theory interpretation 43
2.2.3a Semi-infinite slab approximation 45
2.2.3b Detection 45
2.2.3c Interpretation 47
2.2.3d Simplified interpretation 47
2.2.3e Conclusions 48
2.3 NON-LINEAR RESPONSE CHARACTERISTIC 49
2.3.1 Logarithmic characteristic 50
2.3.2 Self calibrating action 51
2.3.2a Implications 52
2.3.3 Interpretation of resulting signals 52

2.4 MODELLING THE EFFECTS OF MOTION ARTEFACT 53
2.4.1 Model determination 54
2.4.2 Physical investigation 55
2.4.2a Experimental methodology 56
2.4.2b Results 58
2.4.3 A heuristic artefact model 59
2.4.3a Consequences of the model 61
2.4.3b Consequences for pulse oximetry 61
2.5 SUMMARY 63

CHAPTER 3
ARTEFACT REDUCTION 65

3.1 MOTION ARTEFACT EQUALISATION 66
3.1.1 Non-linear equalisation 66
3.1.2 Practical considerations 68
3.1.2a Geometric source positioning 68
3.1.2b Source intensities 68
3.1.2c Direct coupling 69
3.1.2d Spectroscopic measurement 70
3.1.3 Consequences for residual ambient artefact 70
3.1.3a The effects of equalisation 71
3.1.3b Conclusions 72
3.2 A NEW METHOD FOR PULSE OXIMETRY HAVING INSENSITIVITY TO ARTEFACT 72
3.2.1 Methodology 73
3.2.2 A new method for pulse oximetry 74
3.2.2a Separation of optical paths 75
3.2.2b Ratio of ratios 76
3.2.2c Source dependent simplification 77
3.2.2d Conclusions 77
3.2.3 Relationship with the conventional formulation 79
3.2.3a Uniform transmission model 79
3.2.3b Comparison with the conventional formulation 80
3.3 AN ARTEFACT REDUCING PPG SYSTEM 81
3.3.1 Conventional PPG system 82

3.3.1a System overview 82
3.3.1b Optical receiver 84
3.3.1c Sample-and-hold demultiplexers 85
3.3.1d Timing considerations 85
3.3.1e Filtering and amplification 86
3.3.1f ADC and digital output 86
3.3.2 Artefact reduction methodology 87
3.3.3 Choice of technology 88
3.3.3a Dynamic range 88
3.3.3b Quantisation errors 89
3.3.3c Conclusions 91
3.3.4 Analogue implementation 91
3.3.4a Non-linear circuitry 92
3.3.4b Feedback control 93
3.3.4c Feedback circuit analysis 94
3.3.4d Conclusions 95
3.4 SUMMARY 96

CHAPTER 4
EXPERIMENTAL INVESTIGATION 98

4.1 CONFIRMATION OF THEORETICAL RESULTS 99
4.1.1 Self calibrating action 99
4.1.2 Reduction of residual ambient artefact 100
4.1.3 Motion artefact equalisation 101
4.2 QUANTITATIVE ARTEFACT MEASUREMENT 103
4.2.1 Artefact recognition techniques 103
4.2.1a Shape based recognition 104
4.2.1b Verification by additional transducer 105
4.2.1c Spectral techniques 106
4.2.2 Spectral artefact quantification 107
4.2.2a PPG power spectra 109
4.2.2b Artefact quantification 110
4.2.3 Spectral windows 111
4.2.3a Signal isolation 113

4.2.3b Artefact reconstruction 115
4.3 QUANTITATIVE INVESTIGATION OF ARTEFACT REDUCTION 116
4.3.1 Experimental methodology 116
4.3.1a Probe-coupling artefact 117
4.3.1b Complex artefacts 117
4.3.1c Data acquisition 118
4.3.2 Experimental results 118
4.3.2a Distal-phalangic artefact 119
4.3.2b Mid-phalangic artefact 120
4.3.2c Pseudo-random waving artefact 121
4.3.2d Conclusions 122
4.4 SUMMARY 122

CHAPTER 5
CONCLUSIONS \ DISCUSSION 124

5.1 CONCLUSIONS 125
5.2 SUGGESTIONS FOR FUTURE WORK 126
5.2.1 Remote PPG 126
5.2.1a Preliminary investigation 126
5.2.1b Probe design 128
5.2.2 Extension of the model 128
5.2.2a Additive artefact 129
5.2.2b Reduction of additive artefact 129
5.2.2c Wavelength dependent additive artefact 130
5.2.3 Wavelength selection 130
5.2.3a Beer-Lambert law simplification 131
5.2.3b Diffusion theory interpretation 131
5.2.4 Calibrated pulse oximetry 132
5.2.4a Wavelength selection 132
5.2.4b Additional calibration factors 133
5.2.4c Insensitivity to artefact 134

REFERENCES 136

APPENDIX I
ELECTRONIC CIRCUITRY 145

I.1 FRONT-END CIRCUITRY 146

I.1.1 Optical receiver 147

I.1.1a Zero-bias configuration 147

I.1.1b Linear-bias configuration 148

I.1.1c Transimpedance configuration 150

I.1.1d Implementation 151

I.1.2 Circuit timing 151

I.1.2a Multiplexing and demultiplexing 152

I.1.2b Analogue to digital conversion 155

I.1.2c Digital output 157

I.1.2d Software Implementation 159

I.1.3 Implementation details 162

I.1.3a Power supply 162

I.1.3b External connectors 162

I.2 FILTERING AND AMPLIFICATION 163

I.2.1 Pre-processing filters 164

I.2.1a Multiple feedback filter 164

I.2.1b Implementation 164

I.2.2 Amplifying filters 166

I.2.2a Common amplification stage 166

I.2.2b Implementation of common stage 167

I.2.2c Cascaded amplifying filter 168

I.2.3 Complete signal processing 170

I.3 LED DRIVERS 171

I.3.1 Methodology 172

I.3.1a Current control 172

I.3.1b Switching methodology 173

I.3.1c LED connections 174

I.3.2 Implementation 175

I.3.2a Feedback control 176

I.3.2b Switches 177

I.3.2c Power sharing 177

I.3.2d Digital interface 178
I.4 EQUALISATION CIRCUITRY 179
I.4.1 Principle of operation 180
I.4.2 Implementation details 182
I.4.3 Power supply 182

1. INTRODUCTION

Photoplethysmography is a valuable clinical tool for performing non-invasive physiological monitoring and analysis. The recent proliferation in the use of photoplethysmographic devices has been fuelled by the success of pulse oximetry, the dominant clinical application. The sensitivity of these optical measurements to corruption from external dynamics is well known and is emphasised by abundant reports of difficulties encountered in the practical use of pulse oximeters. The complex and unpredictable nature of signal corruption, or artefact, has impeded extant attempts to address the problem, which have achieved only limited success and then to the detriment of generality. It is therefore postulated that true progress can only be made by a fundamental approach that endeavours to incorporate external dynamics into a physical photoplethysmographic model.

1.1 THESIS OVERVIEW

The principal factor limiting the accuracy of photoplethysmography and its derived technologies is signal corruption by artefact, primarily induced by patient movement or ambient lighting. The aim of this thesis is to explore the possibility of reducing artefact corruption of photoplethysmographic signals in real time. This would result in improvement in the accuracy of indices derived from these signals, in addition to widening the applicability of all photoplethysmographic technologies.

This first chapter begins by introducing photoplethysmography and pulse oximetry, and includes a discussion about the origins and effects of signal artefact. Extant commercial techniques for the reduction or suppression of artefact-induced errors in the context of pulse oximetry are then summarised, and the suitability of these techniques for fundamental photoplethysmographic artefact reduction is examined. The author has previously published much of this discussion in the paper:

M. J. Hayes, P. R. Smith, D. M. Barnett, M. D. L. Morgan, S. Singh, D. D. Vara, "Quantitative Investigation of Artefact in Photoplethysmography and Pulse Oximetry for Respiratory Exercise Testing", in *Proceedings CNVD '97 (Computer-aided Noninvasive Vascular Diagnostics)*, Paris, VDI (Verein Deutscher Ingenieure) **263**, pp. 117-124 (1997).

In chapter 2 a novel model is developed that describes both artefact and the photoplethysmographic signal. Whilst physical interpretation of the origin of the physiological signals is explored theoretically, the origins and effects of artefact are investigated experimentally. It is also demonstrated that a measurement non-linearity can aid in the physical interpretation of received signals and can further provide simplifications in deterministic artefact reduction. The author has also previously published the model and analysis presented in this chapter:

M. J. Hayes and P. R. Smith, "Artifact Reduction in Photoplethysmography", Applied Optics, **37**(31), pp. 7437-7446 (1998).

Chapter 3 discusses the implementation of non-linear artefact reduction, with emphasis placed upon both theoretical and practical issues affecting overall system performance. This technique is then used to develop a new method for pulse oximetry having inherent insensitivity to artefact and possessing increased generality. The chapter concludes by giving an overview of the practical implementation of an electronic system that was designed to achieve real-time photoplethysmographic artefact reduction. The methods described in this chapter are both previously published and protected by patent application:

P. R. Smith and M. J. Hayes, "Motion artefact reduction", UK Patent Application No. 9727085 (1997).

P. R. Smith and M. J. Hayes, "Towards Artefact Free Quantitative Photoplethysmography", in *Proceedings CNVD '98 (Computer-aided Noninvasive Vascular Diagnostics)*, Berlin, VDI (1998).

A quantitative investigation into the scope and performance of this approach is undertaken in chapter 4. Investigation of a number of fundamental artefact-inducing experimental situations is performed using the practical implementation developed in the previous chapter. Results from this investigation have also been previously published:

M. J. Hayes and P. R. Smith, "Quantitative evaluation of photoplethysmographic artefact reduction for pulse oximetry", in *Proceedings BiOS '98 (Biomedical Optics)*, Stockholm, ELA / EOS / SPIE (1998).

Chapter 5 briefly reiterates the major achievements of the work and provides concluding remarks. Recommendations are also made for further studies in this area.

The references are collected in a common section following the main body of the report. Appendices are included that give technical details about the electronic designs used to implement the practical artefact-reduction system.

1.2 PHOTOPLETHYSMOGRAPHY

Photoplethysmography is the electro-optic technique for non-invasive monitoring of peripheral blood volume changes, which has found extensive use in the field of biomedical optics. The ease of use, simplicity and non-invasive nature of measurements based upon this successful technique have resulted in the dominance of optical methods as the preferred technology in a number of biomedical monitoring situations. This brief introduction explains the basic principles of operation of photoplethysmographic systems, and highlights the current major clinical applications.

1.2.1 Principle of operation

The term photoplethysmography[1] (PPG), meaning optical plethysmography (the registration of volume change), was originally used to describe the electro-optic technique of measuring the cardio-vascular pulse wave found throughout the human body. This observed pulse wave, now referred to as the photoplethysmographic signal (or *blood volume pulse* as it is known in the USA), is caused by the periodic pulsations of arterial blood within the peripheral vasculature and is discernible by the dynamic optical absorption that this induces in well-perfused peripheral tissue.

1.2.1a Transmission and reflection modes

Observation of the PPG signal can be achieved by illumination of a suitable pulsating vascular bed, for example by transillumination of the fingers or toes. Infrared light is commonly used because it is well absorbed in blood and weakly absorbed in tissue; blood volume changes are therefore observed with reasonable contrast. As the illuminated vascular bed pulsates, the optical path length through it alters and therefore the transmitted light is modulated throughout the cardiac cycle. The monitoring sites for this *transmission mode* PPG are limited to well perfused areas of the body that are transparent enough for the transmitted light to be easily detected, such as the finger tips, toes and earlobes, and in the case of infants the foot or palm of the hand. In *reflection mode* PPG both source and detector are positioned at the skin

surface, with back-scattered light returning from a range of depths within the highly scattering tissue. In the near infrared region light penetrates several centimetres into tissue[2], where the dynamic absorption of the pulsating vascular bed modulates the total reflected light. Reflectance PPG allows monitoring on other peripheral sites such as the forehead, limbs and chest[3].

1.2.1b PPG signal components

Observation of the transmitted or reflected intensity indicates that the pulsatile component accounts for only a very small proportion (1 to 5%) of the total intensity[4]. This results from the fact that a vascular bed may contain only a small amount of blood, which itself experiences only a volumetric change of a few percent with the cardio-vascular pulse wave. The small arterial blood volume changes induce a change in optical path length and modulate the macroscopic absorption of the

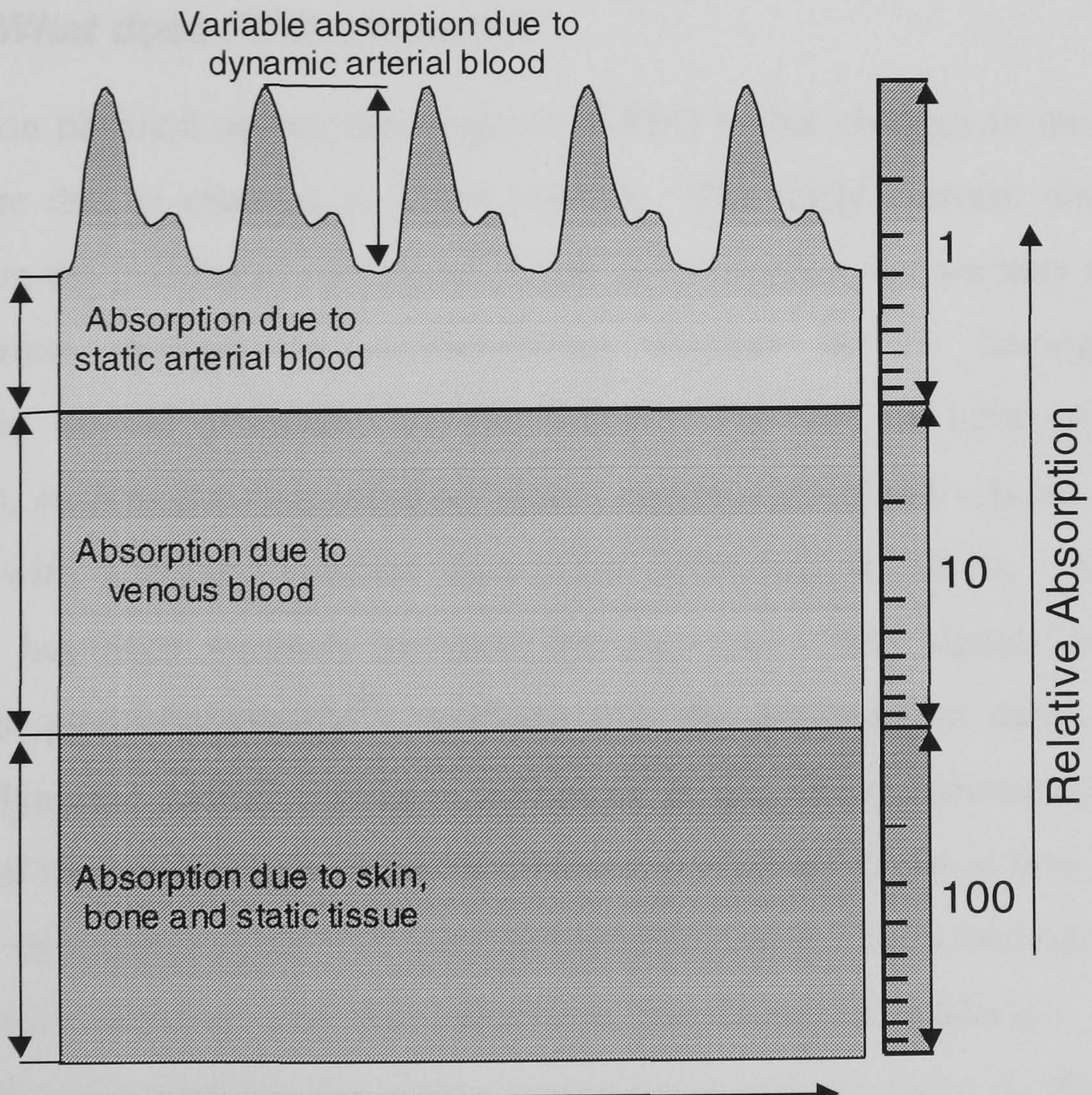


Figure 1.1 – Dynamic absorption of vascular bed

vascular bed (see figure 1.1). The amplitude of these arterial volume pulses has been physiologically interpreted as a measure of the blood supply to the skin[5]. Since blood circulation in the body undergoes a transition from high-pressure arteries to low-pressure veins, much of the non-pulsatile blood will be venous. Although it is difficult to distinguish between static blood absorption and other effects such as absorption by skin, tissue and other anatomical components, the non-pulsatile intensity has found use in comparative venous testing[6-8]. Any electrical representation of the intensity used in a practical PPG realisation will therefore contain an AC, or dynamic, component that represents the arterial pulsations and a much larger DC, or slowly changing quasi-static component, which is indicative of the venous blood volume. The majority of clinical technologies based upon PPG techniques utilise the arterial pulsations to optically isolate absorption by arterial blood, whilst relying on the quasi-static signal component for calibration.

1.2.1c What does PPG measure?

The common physical assumption implicit in PPG is that changes in measured light intensity are due to changes in blood volume. The fairly uniform distribution of blood within the peripheral vasculature leads us to suppose that we may model those blood volume changes by corresponding changes in the homogenous and macroscopic optical properties of the tissue. The use of light of a suitable wavelength, such as the near-infrared region, ensures that blood volume changes are registered with a greater contrast than other physiological effects. A significant correlation has been reported between dynamic (AC) PPG signals and dynamic strain-gauge plethysmography[9], verifying that the predominant cause of the AC signal is dynamic blood volume. Although it has been demonstrated that the amplitude of this arterial PPG signal is related to variations in blood flow and volume throughout the cardiac cycle[10,11], no exact relationship has been formulated between the observed pulsations and the underlying physiological dynamics. The over-simplistic Beer-Lambert law for optical transmission through tissue is often employed to aid physical understanding[12], with broader theoretical applicability afforded by diffusion theory[13,14].

Whilst the received light intensity depends on many factors, both physiological and geometric, appropriate sensor design coupled with compensation for skin absorption by control of the source intensity[15] can result in a quasi-static (DC) PPG signal that is governed to a large extent by the total illuminated blood volume. Dynamic changes in the optical properties of the tissue, such as the arterial pulsations, can further be used to isolate absorption by dynamic blood from absorption by static tissue components. In this way non-invasive optical characterisation of blood may be performed despite the complex optical interactions and the inherent variability in both subjects and geometry.

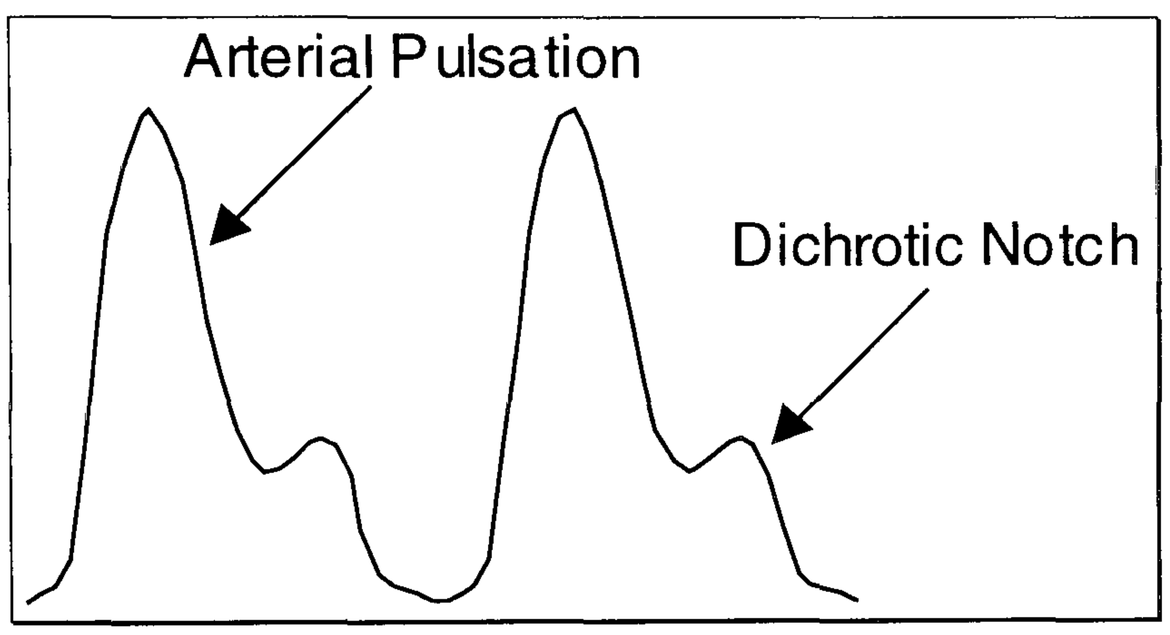


Figure 1.2 – Shape of arterial pulsations

Analysis of the detailed temporal and spectral characteristics of PPG signals can also be used to infer physiological information. For example, the presence of a second, smaller pulse within an arterial pulsation, called a *dichrotic* notch (figure 1.2), can be interpreted as the result of arterial back-pressure when the heart valve closes. Low frequency trends in the quasi-static PPG level and arterial pulsation amplitude have been attributed to the effects of respiration and vasomotion[16].

1.2.2 Major applications

At present, the major application for PPG is the pulse oximeter[17], which utilises a knowledge of the relative absorption of haemoglobin and oxy-haemoglobin to obtain a non-invasive measure of arterial oxygen saturation. Others have used arterial PPG signals to measure physiological variables such as pulse rate[18], respiration rate and vasomotion[16], blood flow[10], arterial blood pressure[19] and viscoelastic properties of

the vessels[20]. Venous PPG signals have been used in a variety of functional venous haemodynamic testing situations[6-8,21], with recent advances in the calibration of these signals[15] spawning increased clinical interest. We examine here two applications of PPG signals, those of pulse oximetry and venous haemodynamic testing. Pulse oximetry not only represents the dominant clinical application of PPG signals, but also serves to demonstrate the use of arterial pulsations to isolate absorption due to arterial blood, enabling a non-invasive spectral blood analysis. Venous haemodynamic testing illustrates the use of the quasi-static (venous) signal component to perform a comparative measure of total illuminated blood volume.

1.2.2a Venous haemodynamic testing

The main established clinical tests that utilise venous PPG signals alone are the muscle-pump test and the vein-occlusion test. Both tests provide useful timing information, with additional information potentially being provided by calibration of the PPG signals[22].

The muscle pump test evaluates the blood displacement in the peripheral venous system under muscle stress conditions, enabling screening for various venous disorders such as varicose veins, deep vein thrombosis and venous dysfunction. The patient performs exercise, such as dorsal extensions or half knee bends, resulting in venous blood being pumped upwards. In a patient with healthy veins the blood cannot flow back because of the venous valves. This results in a detectable change in the quasi-static PPG signal. Following the exercise the veins refill by arterial inflow, with the result that the PPG signal returns to its initial value. If the valves are faulty then the refilling time will be shortened by a time corresponding to the severity of the fault.

The vein-occlusion test is used in the diagnostic investigation of venous obstruction, which may afflict patients with deep vein thrombosis. An occlusion cuff is used to prevent venous outflow, resulting in a PPG detectable increase in peripheral blood volume. After a pre-determined time the cuff is released and the venous outflow can be monitored.

1.2.2b Pulse oximetry

The term *oximeter* was originally coined to describe a continuously transilluminating system[23] designed to measure the oxygen saturation of blood using different wavelengths of light, which operated on a heated earlobe. This methodology relied on the knowledge that haemoglobin and oxy-haemoglobin absorb light to varying degrees as a function of wavelength[24]. Deoxygenated haemoglobin absorbs more light in the red, oxygenated haemoglobin absorbs more light in the infrared (see figure 1.3). Illumination of blood (by either transmission or diffuse reflection modes[25]) with two light sources at different wavelengths thus produces a contrast that is dependent on the relative concentrations of the haemoglobin species, and therefore the oxygen saturation. Consideration of the optical spectra highlights that a choice of wavelengths on opposite sides of the *isosbestic point*, that is the wavelength at which the absorption of the haemoglobin species are identical (around 800 to 815nm), will give the greatest contrast possible. It is usual in commercial systems to use red and infrared sources, which are both spectrally suitable and easily fabricated,

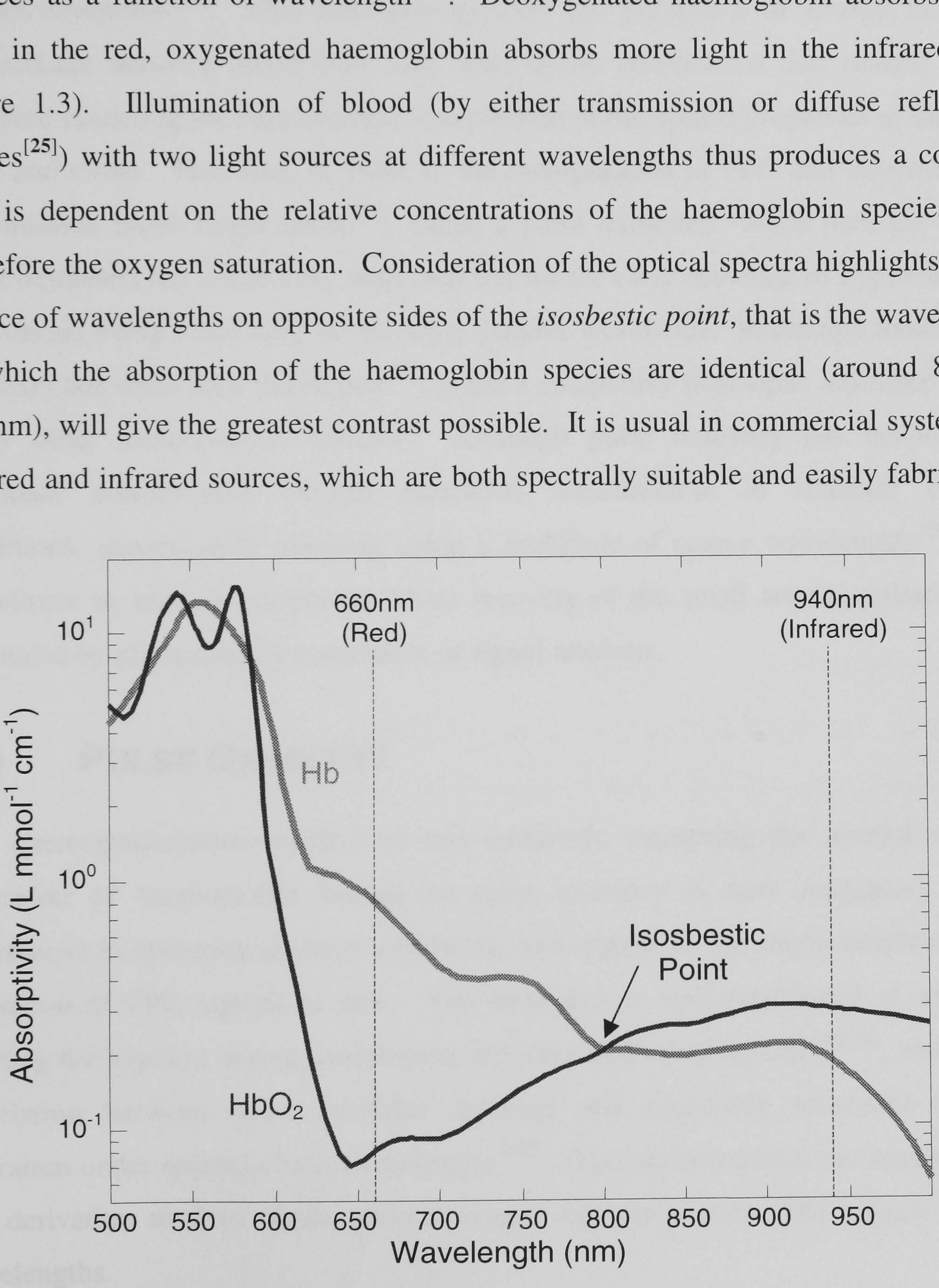


Figure 1.3 – Absorption spectra of haemoglobin species

although recent developments suggest there may be some advantages in using a pair of near-infrared sources[26] either side of the isosbestic point.

An improvement on the earpiece oximeter, which alleviated the requirement for subject heating and aided calibration, was developed when it was discovered that the absorbency ratio of PPG pulsations at two wavelengths changed with the arterial oxygen saturation[12]. This methodology used the pulsations of arterial blood to differentiate between blood flow and other tissue constituents that remain static, therefore rendering the measurement independent of the optical properties of the skin, bone and tissue. This lead, in 1980, to the incorporation of PPG and oximetry in a transmission mode finger probe[17], called a pulse oximeter. Since then the use of pulse oximeters has burgeoned, with their use for the early detection of hypoxia under anaesthesia being mandatory in operating theatres worldwide. Reflection mode pulse oximetry has since been developed[27], as has a completely fibre-optic oximeter[28] and many other technological variants. Although pulse oximetry has become the dominant non-invasive oxygen saturation measurement in standard clinical conditions, non-pulsatile oximetry using a multitude of source wavelengths[29] may sometimes be used in conditions where recovery of the small arterial pulsations is precluded by physiological constraints or signal artefacts.

1.3 PULSE OXIMETRY

The spectrophotometric method of non-invasively measuring the arterial oxygen saturation of haemoglobin known as pulse oximetry is now mandatory under anaesthesia in operating theatres worldwide, and represents the single largest clinical utilisation of PPG signals to date. The technique is well established as an early warning for hypoxia during anaesthesia, recovery and intensive care[30,31], with good correlation between pulse oximeter readings and invasively measured oxygen saturation under normal clinical conditions[3,32]. This section details the assumptions and derivation used to obtain arterial oxygen saturation from PPG signals at two wavelengths.

1.3.1 Principle of operation

Oxygen transport capability, that is the degree to which oxygen is chemically combined with haemoglobin, is referred to as the *functional arterial oxygen saturation*, or SaO_2. This quantity is defined as the ratio of the concentration of oxygen carrying haemoglobin (oxy-haemoglobin HbO_2) to the sum of deoxy-haemoglobin and oxy-haemoglobin ($Hb + HbO_2$) present in the blood stream,

$$SaO_2 = \frac{HbO_2}{HbO_2 + Hb} \times 100\% . \qquad [1.1]$$

SaO_2 is closely related to the *fractional arterial oxygen saturation*, which is the ratio of oxy-haemoglobin to the total haemoglobin of all species (including carboxyl haemoglobin, methaemoglobin etc.). Since the amount of oxygen physically dissolved in blood is small in comparison to the oxygen chemically combined with haemoglobin, the functional or fractional arterial oxygen saturation gives a good measure of the total oxygen content of blood.

The spectroscopic measurement used to distinguish between the haemoglobin species relies on illuminating the vascular bed simultaneously at two wavelengths, λ_1 and λ_2, with the transmitted or reflected intensity due to each illumination wavelength being measured. In practice the sources must be multiplexed so that each source provides illumination for a discrete time period during which the intensity may be measured.

1.3.2 The Beer-Lambert law

The Beer-Lambert model for optical transmission through a turbid medium,

$$I(\lambda) = I_0(\lambda)\exp\left(-\mu_{eff}(\lambda)r\right), \qquad [1.2]$$

couples path length, r, and effective absorbance, $\mu_{eff}(\lambda)$, into a single measure of optical density, and is usually used in pulse oximetry as a basis for theoretical understanding[12]. The constant I_0 is interpreted here as the source intensity, and has wavelength dependency to account for the discrete narrow-bandwidth sources used in pulse oximetry. Whilst the physical applicability of the Beer-Lambert law is questionable[33,34], it has been shown that this approximation to the complex optical interactions that occur in tissue can give a good first order solution[13]. Contemporary pulse oximeters use empirical calibrations that are conceptually based upon a Beer-Lambert formulation. Induced hypoxia experiments are commonly employed to obtain the calibration constants, which give reasonable performance within the bounds of the practical calibration set[35,36]. The primary assumption of the Beer-Lambert formulation is that the optical path through tissue may be separated into a static component and a small dynamic component that depends only on the wavelength-dependent absorbance of the pulsating arterial blood. This can be expressed by decomposition of the macroscopic optical density,

$$\mu_{eff}(\lambda)r = \mu_{blood}(\lambda)d(t) + \mu_{tissue}(\lambda)r\,, \qquad \textbf{[1.3]}$$

where $\mu_{blood}(\lambda)$, $\mu_{tissue}(\lambda)$ are effective absorbances for the blood and tissue respectively, d(t) is a dynamic path length through the pulsating blood (assumed to be identical for all wavelengths) and r is a constant path length through all other anatomical components. The dynamic component is further decomposed by virtue of the absorption difference between oxy-haemoglobin and deoxy-haemoglobin at the illuminating wavelength,

$$\mu_{blood}(\lambda) = S\mu_{HbO_2}(\lambda) + (1-S)\mu_{Hb}(\lambda)\,. \qquad \textbf{[1.4]}$$

S is the proportion of haemoglobin that is oxygenated, in other words the desired arterial oxygen saturation, SaO_2. This allows us to write the received intensity at a

particular wavelength, λ, in terms of the oxygen saturation, the constant effective absorbances and the path lengths,

$$I(t,\lambda) = I_0(\lambda)\exp-\left\{\left(S\mu_{HbO_2}(\lambda)+(1-S)\mu_{Hb}(\lambda)\right)d(t)+\mu_{tissue}(\lambda)r\right\}. \qquad \textbf{[1.5]}$$

It is therefore possible to generate instances of equation [1.5] at different wavelengths and solve these to obtain a value for the oxygen saturation. The number of unknown wavelength dependent terms, and therefore the number of equation instances, can be reduced by a specific solution method that utilises the inherent independence between the static and dynamic intensity components. The following example solutions are conceptually based upon the two predominantly used algorithms, and serve to introduce both common methodology and terminology.

1.3.2a Logarithmic solution

Because of the exponential nature of the Beer-Lambert law, the natural logarithm of the intensity can be used to isolate the multiplicative source constant, $I_0(\lambda)$, and obtain a linear equation in S,

$$\ln I(t,\lambda) = \ln I_0(\lambda) - \left(S\mu_{HbO_2}(\lambda)+(1-S)\mu_{Hb}(\lambda)\right)d(t)+\mu_{tissue}(\lambda)r\,. \qquad \textbf{[1.6]}$$

Since both the source constant and the tissue absorbance are time invariant, we can eliminate them by considering the time derivative of equation [1.6],

$$\frac{d[\ln I(t,\lambda)]}{dt} = -\left(S\mu_{HbO_2}(\lambda)+(1-S)\mu_{Hb}(\lambda)\right)\frac{d[d(t)]}{dt}. \qquad \textbf{[1.7]}$$

The derivative of $d(t)$ is an unknown dynamic but, according to our model, is not wavelength dependent. Thus two instances of equation [1.7] may be generated at the two wavelengths and a ratio constructed to eliminate any dependence on $d(t)$,

$$\frac{\dfrac{d[\ln I(t,\lambda_1)]}{dt}}{\dfrac{d[\ln I(t,\lambda_2)]}{dt}} = \frac{S\mu_{HbO_2}(\lambda_1)+(1-S)\mu_{Hb}(\lambda_1)}{S\mu_{HbO_2}(\lambda_2)+(1-S)\mu_{Hb}(\lambda_2)} = \frac{\mu_{blood}(\lambda_1)}{\mu_{blood}(\lambda_2)} = R\,. \qquad \textbf{[1.8]}$$

The number R, determined from the measured intensities at the two wavelengths, is known as the *ratio of ratios* and is defined as the ratio of the absorbance of blood at the two illuminating wavelengths. Thus the unknown oxygen saturation, S, may be determined from the measured R and a knowledge of the four effective absorbances,

$$S = \frac{R\mu_{Hb}(\lambda_2)-\mu_{Hb}(\lambda_1)}{R\left(\mu_{Hb}(\lambda_2)-\mu_{HbO_2}(\lambda_2)\right)-\left(\mu_{Hb}(\lambda_1)-\mu_{HbO_2}(\lambda_1)\right)} = \frac{SaO_2}{100\%}\,. \qquad \textbf{[1.9]}$$

1.3.2b Linear solution

An alternative solution method[37] directly considers the derivative of equation [1.5],

$$\frac{d[I(t,\lambda)]}{dt} = -I(t,\lambda)\left(S\mu_{HbO_2}(\lambda)+(1-S)\mu_{Hb}(\lambda)\right)\frac{d[d(t)]}{dt}\,. \qquad \textbf{[1.10]}$$

According to the same assumptions as the logarithmic solution method, we can normalise two instances of equation [1.10] with respect to the instantaneous intensities, enabling us to identify a new method of calculating the ratio of ratios,

$$\frac{I(t,\lambda_2)\dfrac{d[I(t,\lambda_1)]}{dt}}{I(t,\lambda_1)\dfrac{d[I(t,\lambda_2)]}{dt}} = \frac{S\mu_{HbO_2}(\lambda_1)+(1-S)\mu_{Hb}(\lambda_1)}{S\mu_{HbO_2}(\lambda_2)+(1-S)\mu_{Hb}(\lambda_2)} = R\,, \qquad \textbf{[1.11]}$$

which no longer requires computation of natural logarithms to calculate R from the measured intensities. This can simplify the processing required in a practical system.

1.3.3 Practical considerations

In a practical realisation, the source multiplexing and analogue-to-digital conversion will form a discrete system. The discrete time samples of the continuous measured intensities are $I(t_1,\lambda)$, $I(t_2,\lambda)$, $I(t_3,\lambda)$ etc, where $t_3 - t_2 = t_2 - t_1$ is the sampling period of the system. Many different methods can be used to estimate R, and therefore the oxygen saturation, from the discrete samples. Whilst R may be averaged over many calculations to obtain a stable result, smaller errors are obtained by averaging the components of R before the ratio is taken[38]. The analysis presented in the following sections develops a best-fit regression calculation of R from the data set of discrete samples, a methodology that is both advantageous in terms of its susceptibility to measurement noise and general enough to illustrate the principles used in many other practical formulations.

1.3.3a Logarithmic solution

It is possible to express equation [1.7] in terms of the discrete sample values, enabling R to be calculated directly from the discrete intensity samples,

$$\frac{d[\ln I(t_1,\lambda)]}{dt} = \frac{\ln I(t_2,\lambda) - \ln I(t_1,\lambda)}{t_2 - t_1} = \frac{\ln\left\{\dfrac{I(t_2,\lambda)}{I(t_1,\lambda)}\right\}}{t_2 - t_1}. \qquad \textbf{[1.12]}$$

Although a sample by sample calculation of R may be useful, it is more practical to consider an averaged value. A better method is to consider the general time periods t and $t+\Delta t$ where Δt is the sampling period,

$$\frac{d[\ln I(t,\lambda)]}{dt} = \frac{\ln\left\{\dfrac{I(t+\Delta t,\lambda)}{I(t,\lambda)}\right\}}{\Delta t}. \qquad \textbf{[1.13]}$$

Thus R may be expressed as the slope of the line $Y(t) = RX(t)$, where

$$R = \frac{\ln\left\{\dfrac{I(t+\Delta t,\lambda_1)}{I(t,\lambda_1)}\right\}}{\ln\left\{\dfrac{I(t+\Delta t,\lambda_2)}{I(t,\lambda_2)}\right\}} = \frac{Y(t)}{X(t)}. \qquad \textbf{[1.14]}$$

If a linear regression technique is applied to the data set of $X(t)$ and $Y(t)$ then a best fit value for R may be obtained. This is a particularly stable method for calculating R from many instantaneous measurements, which has an inherent reduction of noise due to outlying data points.

1.3.3b Linear solution

We now describe the received intensity in terms of the sum of a static and dynamic component,

$$I(t,\lambda) = \overline{I(\lambda)} + I_{AC}(t,\lambda), \qquad \textbf{[1.15]}$$

where $\overline{I(\lambda)}$ is the average, or DC, value of the measured intensity and $I_{AC}(t,\lambda)$ is the instantaneous change from the average value, or the AC. Comparison of equations [1.8] and [1.11] indicates that in order to estimate R we now simply require the time derivative of the intensity, without first performing a logarithmic transformation. Because, by definition, the DC intensity is static, the derivative of equation [1.15] is simply the derivative of the AC intensity. This implies that we may separate the dynamic and static signal components by electronic manipulation and write the value R in terms of the AC signals and their instantaneous counterparts,

$$R = \frac{I(t,\lambda_2)\dfrac{d\left[I_{AC}(t,\lambda_1)\right]}{dt}}{I(t,\lambda_1)\dfrac{d\left[I_{AC}(t,\lambda_2)\right]}{dt}} = \frac{I(t,\lambda_2)d\left[I_{AC}(t,\lambda_1)\right]}{I(t,\lambda_1)d\left[I_{AC}(t,\lambda_2)\right]}. \qquad \textbf{[1.16]}$$

This methodology offers practical advantages, because it introduces the possibility of applying enhanced gain to the relatively small AC signal, therefore decreasing the

necessary precision of the analogue-to-digital converter. If we again consider discrete sampled values, this time from both the instantaneous intensity signals and their AC counterparts, then we can again express R in terms of the best fit to a straight line,

$$R = \frac{\dfrac{I_{AC}(t+\Delta t,\lambda_1)-I_{AC}(t,\lambda_1)}{I(t,\lambda_1)}}{\dfrac{I_{AC}(t+\Delta t,\lambda_2)-I_{AC}(t,\lambda_2)}{I(t,\lambda_2)}} = \frac{Y(t)}{X(t)}. \qquad \textbf{[1.17]}$$

A variation on this technique can be performed by making the assumption that the two intensity signals are synchronous and taking a peak measurement from each signal. The peak measurements are then normalised with respect to the average values, creating *normalised pulsatile absorbances*, the ratio of which is equal to R. Although this technique is equivalent to the calculation of R at any pair of points in time, it simplifies practical implementation and offers the possibility of an all-analogue oximeter.

1.3.4 Sensing periods

Simultaneous illumination of the vascular bed at two source wavelengths is achieved by time-division multiplexing of the light sources, with the use of a single receiver. Each source is illuminated for a finite time period, during which the output from the receiver is sampled and stored. Provided that the multiplexing frequency is high enough (the time periods are short enough), then the resulting signals are approximately equivalent to those that would be obtained by distinct and simultaneous illumination due to each source. Although Nyquist's theorem dictates that the bandwidth of the resulting signals will be half the multiplexing frequency, it is usual to use much higher switching frequencies in order to maximise system performance. This methodology not only simplifies the probe design and equalises the physical path lengths for each source, but also eliminates the possibility of receiver mismatch between the channels.

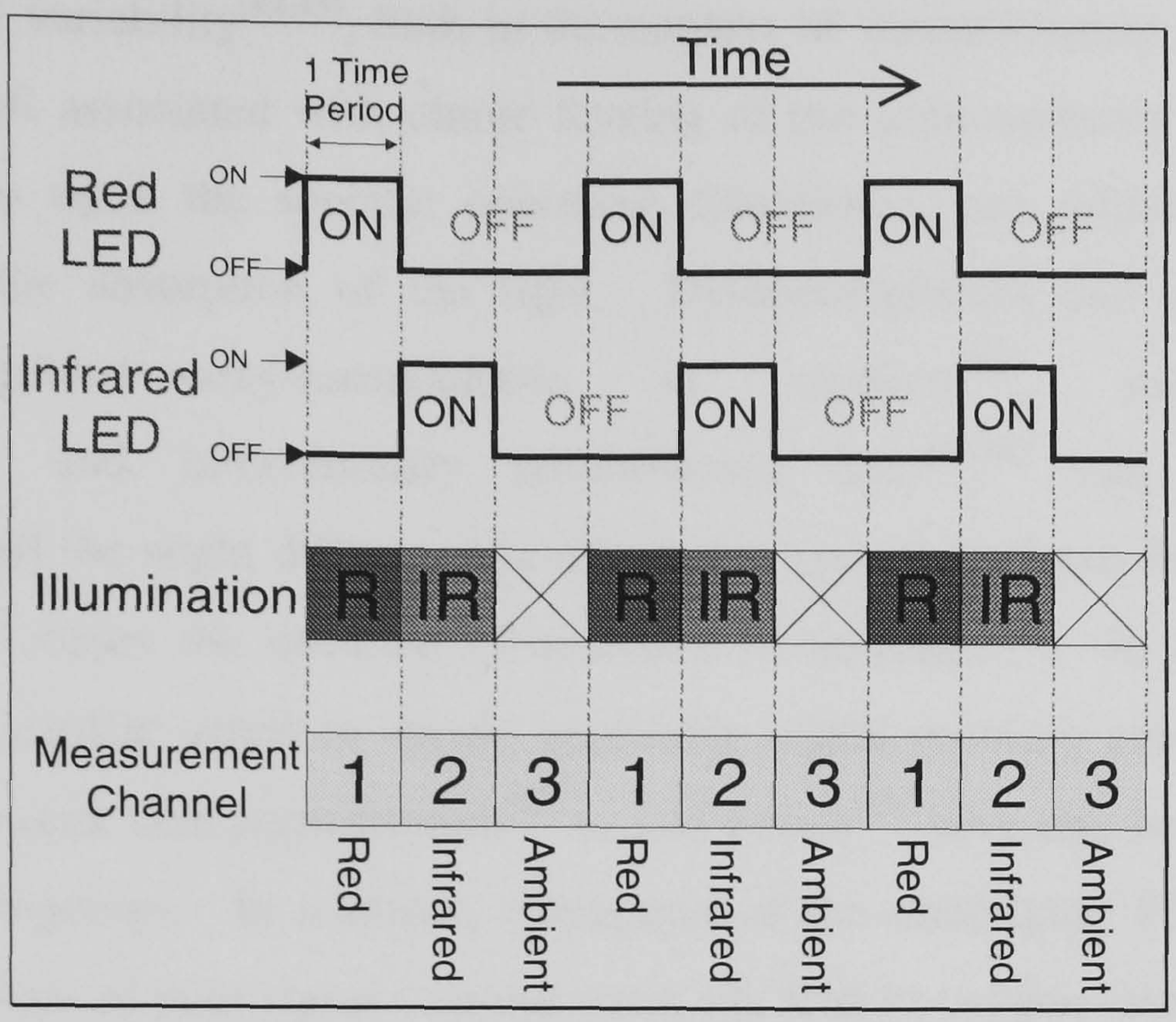


Figure 1.4 – Dual wavelength multiplexing scheme

In a practical system it is necessary to have three sensing periods rather than two. This is so that the ambient lighting effects may be sampled during the third period when both sources are extinguished. The ambient samples may then be subtracted from the intensity samples to remove any dependence on the ambient light (see figure 1.4). The success of this technique depends primarily on the ambient intensity and on the multiplexing speed relative to the modulation frequency of the ambient light. The quantitative effectiveness of this ambient subtraction will be considered in more detail in subsequent sections.

1.3.5 Calibration

Once the ratio of ratios, R, has been estimated using one of the methods outlined above, equation [1.9] may be used to calculate the oxygen saturation. In practice the four effective absorbances will not be derived theoretically, but will be based upon an empirical calibration employing induced hypoxia experiments[33]. Apart from multiple scattering effects (which may be dominated by the haematocrit level[39,40]) and the inapplicability of the Beer-Lambert law[33,34], several factors may influence the practical accuracy of the calibrated measurement in a clinical environment[35,41].

Source spectral variability[42,43], both in the context of manufacturing spread and in the spectral shift associated with ohmic heating of the semiconductor junction[44,39] (which depends upon the specific operating conditions), can produce significant alterations in the absorption of the light. Dyshaemoglobins and dyes, such as increased carbonmonoxy-haemoglobin in smokers[45], methaemoglobin deficiency[42,45] and intravenously administered dyes[42,30] can all produce inaccuracies, and the slight difference in absorption spectra between foetal and adult haemoglobin[46] raises the question of accuracy in neonates[47]. In low perfusion states, such as cardiac arrest or shock, unreliable signal recovery can cause severe errors[30]. Very dark skin pigmentation[48] or nail polish[49] have also been reported to interfere with readings. In addition, corruption of the underlying PPG signals by *artefact*, or periods of poor signal to noise ratio, can lead to oxygen saturation outputs that are seriously in error[53,54]. The effects of signal artefact will be considered in detail in subsequent sections.

These considerations have led to the predominance of empirical calibration techniques, often employing additional calibration factors, and to scepticism about the quantitative accuracy of the measurement when clinical conditions stray from the expected normal, when possibly most accuracy is needed[35].

1.4 LIMITATIONS OF THE PPG MODEL

The conventional PPG model assumes a stable closed system in which the only dynamics are the physiological changes under measurement, such as the arterial pulsations or venous volume changes. This assumption can break down in practice when geometric, physiological or optical dynamics make it impossible to separate the desired measurement from the background noise. Within the context of this model, any measured dynamic that is not due to the desired measurement parameter is termed artefact, and is indistinguishable from the desired PPG information.

1.4.1 Artefact

The term artefact encompasses any spurious signal component that results in poor PPG signal-to-noise ratio. Corruption of the PPG signal can arise from inadvertent measurement of ambient light (ambient artefact) and from voluntary or involuntary subject movement (motion artefact). These artefacts give rise to errors in interpretation of the PPG signals for technologies such as pulse oximetry[50].

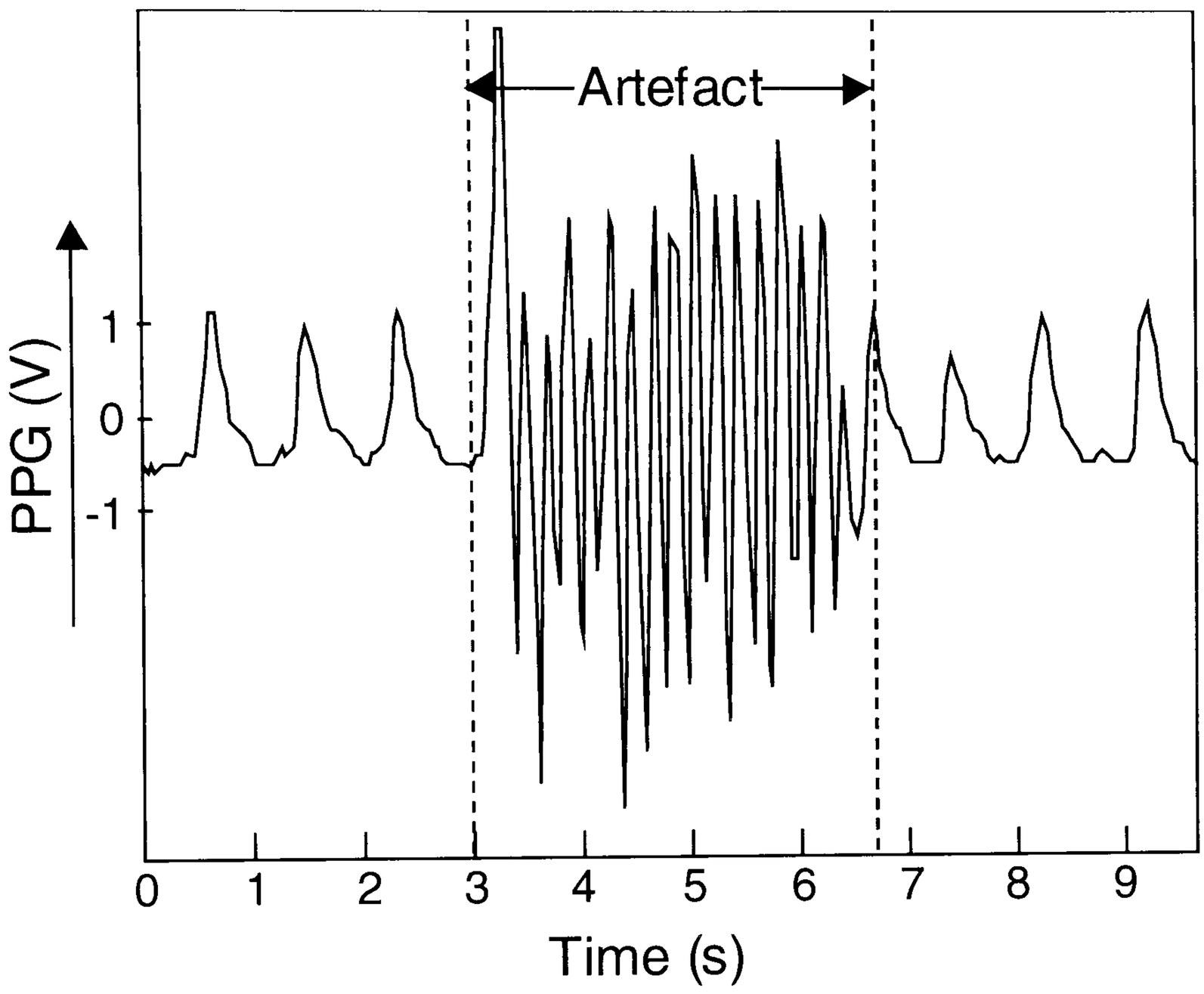


Figure 1.5 – Artefact corrupted arterial PPG signal

The relatively small size of the arterial pulsations means that artefacts are most destructive to the AC signal (figure 1.5, between 3 and 7 seconds), although larger artefacts can also degrade the quasi-static signal component. In addition, the spectral and temporal characteristics of the arterial pulsations may be similar to those of artefacts encountered in practice, increasing the difficulty of isolating or identifying valid pulsations.

Removal of signal artefacts cannot be easily performed by a signal processing or filtering method, because of several key physical features of the observed artefact:

- The artefact may be several orders of magnitude larger than the desired signal, creating a severe signal to noise ratio (SNR) problem for any signal processing scheme.

- In the general case, there is no *a-priori* knowledge, explicit or implied, which can be used to separate artefact from the PPG signal, either temporally or spectrally. This precludes the use of filter based signal processing methods for the removal of artefact.

- The artefact and PPG signals can often be observed to be statistically dependent, a condition that impedes attempts at more sophisticated signal processing methods such as correlation cancellation techniques.

In addition, the real-time nature of PPG systems and the physiological variability encountered between subjects serve to further compound the problem. Although attempts have been made to remove artefact using signal processing techniques, it is only by investigating the origin of artefact that we can progress towards truly artefact free PPG.

1.4.1a Motion artefact

Classes of signal corruption known as *motion artefacts* are the primary sources of unwanted dynamics; so named because they originate from the consequences of voluntary or involuntary patient movement. Motion artefact encompasses corruption of PPG signals due to mechanical forces giving rise to changes in:

- The optical coupling between probe and patient,
- Physiology of the patient,
- Optical properties of the tissue due to geometric realignment,
- Complex combinations of all these effects.

The consequences of a typical patient movement may cause severe artefact that is partially due to all the above mentioned types, with resultant relative movement between patient and probe altering the optical coupling simultaneously with redistribution of venous blood following limb movement.

Apparent changes in venous blood volume may be caused by physiological effects (e.g. exercise or temperature), gravity (e.g. raising a limb or bending down) or pressure (on the probe or the body) and other mechanical forces. As long as these changes take place on time scales greater than a few heart cycles, a conventional arterial PPG implementation will be unaffected, as the methods are usually insensitive to quasi-static changes in the non-pulsatile tissue characteristics. Dynamic venous volumes of a similar time scale to that of the arterial pulsations will, however, cause misinterpretation by technologies such as pulse oximetry. Since fundamental PPG studies in other contexts relate to both arterial and venous blood-volume changes, there is therefore a clear distinction to be drawn between the nature of motion artefact in the context of pulse oximetry and that of PPG. A pulse oximeter considers artefact to be all dynamic components that do not originate from the arterial blood pulsations, whilst a PPG system considers artefact to be all dynamic components that do not originate from changes in blood volume. This observation is important in attempting to precisely define motion artefact. We shall consider only the general case and exclude these non-pulsatile physiological dynamics from our definition. We can then say that motion artefact encompasses all measured dynamics that arise from voluntary or involuntary patient movement that cannot be attributed to effects of physiological origin.

1.4.1b Ambient artefact

The inadvertent measurement of ambient light, or *ambient artefact*, is not a theoretical problem in the recovery of PPG signals. An independent measure of the ambient lighting conditions may be obtained with the use of an additional time period in the multiplexing scheme when all sources are extinguished (see section 1.3.4). The measured ambient signal can then be subtracted from the desired signals during

periods of illumination. A number of clinical publications, however, suggest that ambient light is a problem in practical pulse oximetry systems, mainly due to the severity of ambient artefact experienced in the operating theatre[49-51]. We now examine in more detail the practical limitations which result in non-ideal ambient subtraction, and attempt to quantify the remaining, or *residual*, ambient artefact.

The dynamic range of the intensity receiver will ultimately limit ambient subtraction performance, but only in the case of very high intensity ambient illumination. In this case the receiver will saturate and no signal will be provided. Of more interest here are the effects of a finite switching time for the electronic multiplexing scheme, giving rise to finite residual ambient artefacts that appear as additive noise on the PPG signals. This phenomenon occurs because ambient light is rarely constant; most artificial light is modulated at the mains supply frequency for incandescent bulbs or twice the supply frequency for fluorescent tubes. In addition, passing obstruction or varying cloud cover may modulate apparently constant sunlight, whilst many forms of clinical illumination such as xenon lamps or fibre-optic light sources are directly modulated. This dynamic nature implies that in any discrete-time ambient subtraction scheme, there will be a finite error due to any change that has occurred in the ambient light between the time it was sampled and the time that sample is used for subtraction. We consider a signal due to ambient light, $v(t)$, and a signal that is due to both illumination and ambient light, $v'(t)+v(t)$. The desired illumination signal, $v'(t)$, is simply the difference between the two sampled signals, as expected. If we now build in the effects of a practical discrete-time system, we must sample the ambient signal, $v(t)$, during one time period and the illumination and ambient signal during the next time period, $v(t+T_s)+v'(t+T_s)$, where T_s is the sampling interval between multiplexing periods. Subtraction of the signals now gives

$$v'(t+T_s)+\big(v(t+T_S)-v(t)\big)=v'(t+T_s)+T_s\,\frac{dv(t)}{dt}, \qquad \textbf{[1.18]}$$

which is the desired signal, $v'(t+T_S)$, plus a residual ambient term that is not completely removed by the subtraction. The size of the residual ambient artefact thus depends on the rate of change of ambient light and the multiplexing speed.

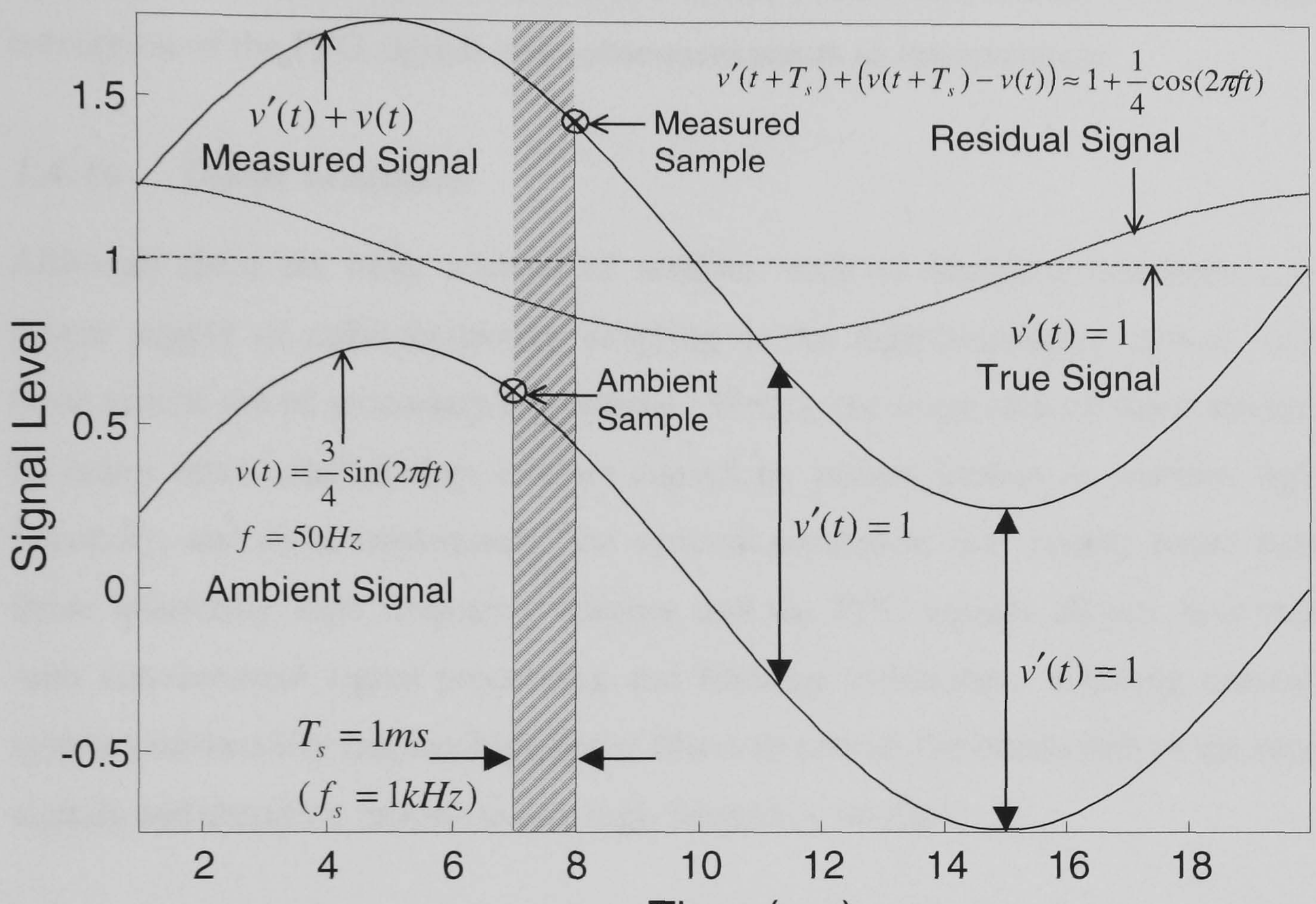


Figure 1.6 – Residual ambient artefact example

To illustrate this dependency, we invoke the example of a sinusoidal ambient signal, such as that caused by incandescent bulbs or fluorescent tubes. Equation [1.18] predicts that the resultant residual artefact will also be a sinusoid, with a magnitude that is $2\pi f / f_S$ of the original signal, where $f_S = 1/T_S$ and f is the frequency of the ambient signal (see figure 1.6). Thus a 50Hz (UK supply frequency) ambient signal from an incandescent bulb will be attenuated by a factor of just 3 (-5dB) when using a 1KHz multiplexing frequency, or by a factor of 30 (-15dB) with 10KHz switching. The use of sensitive optical receivers such as wide area PiN diodes limits the possible multiplexing speeds in practical realisations to a few KHz.

Filters following the ambient subtraction limit the bandwidth of the signals to expected PPG values, enabling us to approximate any subsequent operations as

continuous. They also serve to further attenuate any residual ambient artefact that has a modulation frequency in excess of the expected PPG bandwidth. Nevertheless, with high intensity and high frequency artificial ambient sources the residual artefact may approach or even exceed the magnitude of the small arterial pulsations, resulting in corruption of the PPG signals and subsequent errors in interpretation.

1.4.1c Other artefacts

Although there are other sources of artefact, such as electrical interference in the power supply or radio-frequency coupling to the high-impedance optical receiver, these effects are of secondary importance. Firstly, the magnitude of these effects may be many times smaller than artefact caused by patient motion or ambient lighting. Secondly, and more importantly, the spectral separation that usually exists between these inherently high frequency sources and the PPG signals allows their removal with conventional signal processing and filtering techniques. Existing commercial systems universally employ high-order filters to restrict the bandwidth of the received signals and therefore minimise any high frequency artefact.

1.4.2 The clinical problem of PPG artefact

Owing to its dominance as the major clinical utilisation of PPG signals, it is in pulse oximetry that the most significant end-user problems of PPG artefact are found. This is compounded by the fact that pulse oximeters use the smaller arterial pulsations from the total PPG signal, which are inherently more susceptible to artefact. The primary clinical perception of the problem of artefact is the occurrence of false alarm conditions, which are caused by artefact induced interpretation errors so severe that they falsely imply a clinical emergency. However, transient or less severe artefact corruption of the PPG signals can also cause significant errors in pulse oximeter output that can easily go unnoticed in clinical practice. In addition, the averaging and smoothing algorithms used inside most commercial oximeters have the effect of incorporating false measurements (due to artefact) into a running total[53], making it very difficult to identify and discard incorrect readings.

1.4.2a Corruption of pulse oximetry output

Any underlying artefact corruption of the PPG signals used in a pulse oximeter can be observed to corrupt the outputs[53,54], both pulse rate and oxygen saturation. A high proportion of false alarms in a clinical environment have also been attributed to artefact[55], accounting for the predominant clinical perception of the problem. Since the pulse rate output is easily verified independently, it may be used to determine whether continuous PPG signal corruption has occurred, indicating that the oxygen saturation output should be discarded[53,56]. The inaccuracies introduced by incorporation of transient errors into an averaged value are, however, more difficult to identify or clinically quantify. Patient motion, and therefore implicitly motion artefact, has been clinically demonstrated to cause errors in pulse oximeter output[50,57], impeding the success of ambulatory studies and exercise testing[52,58]. Ambient lighting has also been reported to cause errors in a clinical environment; the most severe cases being due to high intensity or quickly modulated artificial sources[51,59], with even infrared heat lamps[60] and natural sunlight[61] reported to cause significant errors. Although most of the published literature only reports the effects of artefact on the accuracy of pulse oximeter output, it is PPG signal corruption that ultimately leads to these errors.

1.4.2b Applicability of the PPG model

Whilst artefact is frequently described as noise degrading a clean PPG signal, the origins of artefact are as complex and interesting as the origins of the PPG signals themselves. The conventional PPG model does not attempt to include artefact-inducing effects in its physical description. This omission is the primary obstacle to the application of PPG technologies to many possible future measurements, where artefact inherently dominates received signal levels. For example, implementation of non-contact, or remote, PPG could enhance measurements such as foetal or internal monitoring, were it not for the problems of distinguishing the true PPG signals from the multitude of minute motion artefacts. The accuracy of foetal and ambulatory studies is also degraded by artefacts, as is the monitoring of children and exercising patients[52]. It is only by attempting to understand the origin of these effects that we

can widen the applicability of both PPG and derived technologies such as pulse oximetry.

1.5 EXTANT TECHNIQUES FOR ARTEFACT REDUCTION IN PULSE OXIMETRY

It is because of the clinical problems presented by false alarms and general output inaccuracies that most of the attempts to solve the artefact problem are to be found in the context of pulse oximetry. Although most of these solutions concentrate on the reduction of false alarms in a clinical environment, the diverse spectrum of methodologies spawned by this commercial problem warrant further investigation. There is a lack of purely academic literature available on this subject, with the bulk of extant solutions published only in the patent literature.

Most existing attempts to reduce artefact have concentrated on the problem of motion artefact; assuming that reported ambient artefact is due to unrealistically extreme conditions that should not occur in normal practice. Even so, amplitude demodulation at the receiver has been used to enable a higher switching speed[62], which could in principle be used to reduce residual ambient artefact. There have been a number of commercial attempts at tackling the problem of motion artefact in the context of pulse oximetry, which can be said to fall into four categories:

1. Correlation cancellation or signal processing solutions[63-66].

2. Feature based recognition of corrupt pulses[67-70].

3. Cancellation via control of an additional light source[71].

4. Inversion of an empirical artefact model[72].

Both methods 1 and 3 ideally use an independent measurement of artefact supplied by another transducer (e.g. piezo or optical), making the assumption that artefact is a linear addition to the pulsatile signal. It will be shown later that observation of the PPG signal under typical artefact producing forces casts doubt on this hypothesis.

Method 2 is an engineering solution to the end user problem of artefact, as it provides a method of output suppression in the presence of artefact. Method 4 utilises a more general multiplicative model to compensate directly for motion artefact in a modified formulation of the oxygen saturation algorithm.

1.5.1 Correlation cancellation or signal processing

Adaptive noise-cancellation techniques can be used to remove noise from a signal, provided that some reference signal is available that is well correlated with the noise and not the signal[73]. The most obvious method to obtain the necessary reference signal for motion artefact reduction is to independently measure the artefact using some form of movement sensor, such as a piezoelectric film[63], or another light source of a wavelength which is substantially unaffected by dynamic blood absorption[64]. Another method[65] utilises a second signal-processor to obtain the reference signal from two or more intensity signals. The assumption is that the desired portions of each signal will be correlated, as will the artefact components, but that the two types of component will be essentially uncorrelated. Alternatively, constant oxygen saturation may be assumed and instantaneous variations calculated from the intensity samples used to produce a correlated noise signal[65]. Correlation cancellation, like all signal-processing techniques, models the motion artefact as an additive noise signal. Although this model may be over simplistic there is some clinical evidence as to the success of these techniques[74,75], which have achieved reasonable results in the context of alarm reduction in pulse oximetry.

Another signal processing[66] method uses a selective frequency bandpass filter that is adapted to have its centre frequency at the fundamental frequency of the PPG signal, therefore removing any artefacts which do not lie in a relatively small spectral region. This technique is obviously only successful when the artefact is spectrally separate from the desired PPG signals, a condition that is unrealistic in a large proportion of clinical situations.

All signal-processing methods, correlation-cancellation included, suffer from a lack of generality that is imposed by the requirement of stringent *a-priori* definition of

which signal components are desired information and which are spurious noise. This specialisation makes it difficult to apply a single technique to a large number of clinical situations. For example, the venous PPG signal would almost certainly be removed by the majority of techniques designed for application to pulse oximetry, having unknown dynamics and being largely uncorrelated with the arterial signal. In addition, the underlying assumption of any signal processing approach is that artefact may be modelled as a noise term that is a linear addition to the PPG signal. This simplistic model would imply that any received intensity from the arterial pulsations would be unaffected by the underlying motion. It will be demonstrated later that observation of PPG signals under typical artefact producing conditions casts serious doubt on the validity of this model. This questions both the accuracy and physiological relevance of PPG signals that have been processed in a manner dependent on these assumptions. Finally, the complexity of implementation required for the majority of these techniques necessitates the use of some form of digital processing. The extreme signal-to-noise problem of artefact in PPG dictates that any digital processing solution may not be able to completely remove artefact because of the quantisation errors introduced by the sampling process.

1.5.2 Feature based recognition of corrupt pulses

It is possible to deduce some *a-priori* knowledge about the characteristics of expected PPG pulsations, such as the spectral information that is commonly used to limit the bandwidth of the signals. An extension of this idea[67] models salient temporal features of an expected PPG signal and rejects received arterial pulsations that do not conform to its predetermined ideal. Whilst this system does not reduce artefact on the PPG signals themselves, the rejection of artefact corrupted pulses has been reported to significantly reduce the number of pulse oximetry false alarms in a clinical environment[76]. A related method[68] uses a second pulse oximeter to provide comparative recognition of corrupt signals, assuming that the PPG signals and local oxygen saturation are invariant with sensor location. Alternatively, external validation of the device estimation of pulse rate can be used as an indication of periods of artefact corruption[53,56]. The practical implementation of this technique is

hampered by the degree of averaging used to obtain the pulse rate output, and therefore the limited temporal resolution of artefact recognition.

In another methodology, three wavelengths of light may be utilised, with the oxygen saturation being determined simultaneously from two pairs of sources[69]. Provided that any artefact affects all signals in a similar manner, then interpretation of artefact corrupted signals will reveal a difference between the pair of calculated saturation values; indicating that the results should be ignored. Linear interpolation and rate of change techniques, along with selective frequency filtering, have also been used to compensate the detected optical signal by using the filtered or interpolated signal to estimate the oxygen saturation during transient conditions[70].

The nature of artefact recognition by classification of corrupt pulses or invalid interpretations fails to address the fundamental problem of output errors due to artefact, and therefore cannot improve output accuracy during periods of slight artefact. This is because partially corrupt pulses and their interpretation will still be incorporated into a time-averaged result. In addition, these techniques will fail to provide any output at all during extended periods of artefact, such as ambulatory studies or exercise testing, when all pulses will register as corrupt. The general methodology of artefact recognition is an engineering solution to the end-user problem of artefact, and as such the performance will be determined by the scope and accuracy of *a-priori* knowledge used to identify corrupt data. The methods are unsuitable for application to fundamental PPG artefact reduction, because they cannot provide clean data in the presence of artefact and lack the generality required for the wide range of clinical PPG applications.

1.5.3 Cancellation via control of an additional light source

It has been suggested[71] that motion artefact may be suppressed by adaptive control of a pair of coincident source intensities, such that the resulting average, or DC, levels of intensity received from each source are identical. The implicit assumption is that motion artefact is manifested as an additional intensity that is synchronous for all wavelengths and varies only in size. Observation of practical PPG systems suggests

that the size of the additive motion artefact would then be proportional to the DC intensity, with the result that signals whose DC levels are identical would contain identical artefact. Subtraction of two such signals would therefore result in an arterial signal that is independent of the motion artefact term. Because the subtraction mixes information from the pair of wavelengths used, three wavelengths would then be required to reformulate the Beer-Lambert law derivation of pulse oximetry, with subsequent calibration requiring tissue optical parameters at all three wavelengths.

Although there is no direct evidence to date as to the success of this technique, it is likely that a practical implementation could achieve limited artefact reduction. The assumption that motion artefact is additive may be over simplistic, because it implies that any received intensity from the arterial pulsations will not be modified in any way by the underlying motion. However, a more general model could be linearised about the common DC point, with the result that the subtraction process could still reduce sufficiently small artefacts. In addition, the subtraction process will effectively remove any residual ambient artefact, which would be identical on both signals. It is likely, however, that the simplistic artefact model will preclude the use of this technique for reduction of all but the least severe artefacts, where dynamic modification of received light from the actual arterial pulsations would be negligible. In addition, subsequent analysis of the artefact-reduced PPG signals may also be flawed because of this assumption, requiring further investigation before these signals could be used in a practical three-wavelength pulse oximeter.

1.5.4 Inversion of an empirical artefact model

An artefact model that is more general than the additive model described in previous sections can be formulated by assuming that any motion artefact modulates the received intensity in an identical manner for all signals[72]. This then acknowledges that the artefact must in some way affect the received arterial pulsations; in this model all signal components are affected in the same way. It is then possible to rework the oxygen saturation derivation in such a way that the multiplicative motion artefact term is eliminated in the final calibration. The solution method utilises inversion of a

matrix of equations, which ideally requires three wavelengths for unique solution. The modified oxygen saturation algorithm then uses tissue dependent parameters at all three illuminating wavelengths.

The theoretical success of this technique will be entirely dependent on the accuracy of the empirical artefact model used, for which no experimental or theoretical justification is currently available in the open literature. Practically, it is likely that the greatest limit on performance will be the restriction of implementation to digital processing, and more specifically processing in conjunction with the conventional calibration scheme. The inherently wide dynamic range of PPG signals, together with the severity of possible signal-to-noise ratios, suggests that this technique may only be successful under conditions of slight artefact. Severe artefact might still obscure the delicate arterial pulsations, both in terms of the previously mentioned quantisation errors inherent in analogue-to–digital conversion, and by dominating the available precision of internal calculations. In addition, any methodology that operates by modification of the pulse oximetry calibration algorithm is obviously unsuited to fundamental PPG studies, where it is the detailed PPG signal itself that is of interest.

1.5.5 Conclusions

Although there have been many attempts to address the problem of artefact in the context of pulse oximetry, no single method to date stands out as a candidate for application to fundamental PPG studies. Figure 1.7 shows the classification of these attempts by their conceptual origin, with methods that have been claimed as partially successful being further categorised by their limitations. The limitation categories are specificity to pulse oximetry, restriction to artefact recognition not reduction and the use of an additive artefact model. These limitations illustrate the primary obstacles to the application of these techniques to fundamental PPG studies. Further categorisation could include the nature of a-priori assumptions, restriction of implementation to a digital processing solution and the level of generality.

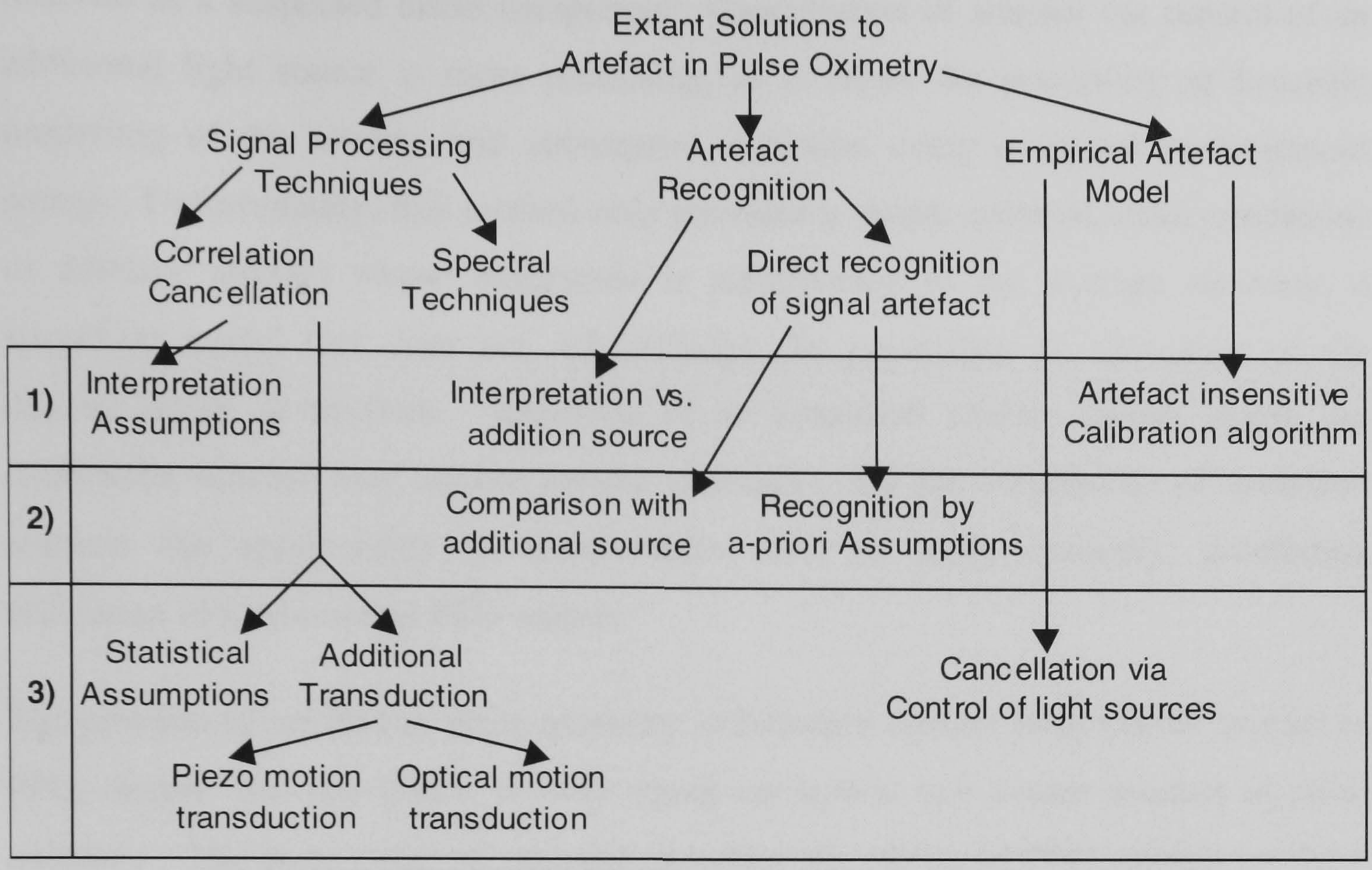


Figure 1.7 – Extant techniques for artefact reduction in pulse oximetry

Recognition of artefact corruption in pulse oximeters may prevent a significant proportion of clinical false alarms, therefore lessening the end-user problem of artefact, but the technique can never be used to provide clean PPG signals in the presence of artefact. Signal processing techniques provide the possibility of processing artefact corrupted signals to remove the undesired signal components, modelling the artefact as an additional noise term. Excluding concerns about the physical relevance of such techniques, their performance will be primarily dependent upon the accuracy and applicability of the underlying *a-priori* assumptions and definitions of what comprises a clean signal and what is noise. Whilst these ideas have had some success in removing unexpected signal components in a pulse oximeter, the lack of generality makes it difficult to apply them to fundamental PPG studies, where the detailed structure of the signals is not known in advance. In addition the venous signal, which can be of interest in PPG studies, may be uncorrelated with the arterial signal and of unknown dynamics, resulting in its

removal as a suspected noise component. Cancellation of artefact via control of an additional light source is more promising, as it offers the possibility of heuristic modelling of the artefact and subsequent inversion using a second measurement source. Unfortunately, this method only provides a simple solution under conditions of additive artefact whose magnitude is proportional to the average intensity; a simplistic model that does not acknowledge the possibility of corruption of the desired signal component. Inversion of an empirical artefact model within the calibration scheme may enable greater generality, but the complexity of inversion restricts the applicability to technologies such as pulse oximetry, precluding utilisation in fundamental PPG studies.

The problem of artefact in pulse oximetry is therefore distinct from that of artefact in PPG, despite the fact that it is PPG signal corruption that causes artefact in pulse oximetry. The wide range of attempts to reduce the effects of PPG artefact on pulse oximeter output not only demonstrate the degree of motivation to solve the problem, but also highlight the difficulty of obtaining a general solution by virtue of their limited degree of success. It is likely that a combination of these contemporary artefact recognition or suppression algorithms could lessen the occurrence of pulse oximeter false alarms, one of the primary motivations for solution of this problem. Recovery of uncorrupted PPG signals in the presence of severe artefact is, however, beyond the scope of these methodologies. It is only by fundamental artefact reduction that the applicability and accuracy of all PPG technologies, pulse oximetry included, may be significantly advanced.

1.6 SUMMARY

The non-invasive clinical monitoring technique of photoplethysmography has been introduced, both within the context of common clinical utilisation and by consideration of the physical origin of the signals. Particular emphasis has been placed upon description of non-invasive arterial oxygen saturation monitoring by pulse oximetry, the biggest single application for PPG signals to date. The prevalent model and assumptions implicit in pulse oximetry have been demonstrated by a

simple Beer-Lambert law derivation, which has served to highlight the common terminology, solution techniques and practical problems of calibration. The prevalence of empirical calibration in contemporary systems has been justified, stressing the desire for future algorithmic generalisation.

The problem of artefact in photoplethysmography has been introduced as a deficiency in the conventional model, with the primary classes of artefact being identified as originating from patient movement and ambient lighting. A detailed examination of the origins of these artefact classes has enabled quantification of the severity of ambient artefact and classification of movement artefact for fundamental PPG studies and applied technologies such as pulse oximetry. It has further been highlighted that many problems in the clinical use of pulse oximetry can be observed to originate from artefact corruption of the underlying PPG signals, resulting in false alarm conditions, incorrect or inaccurate oxygen saturation output and a restriction on the applicability of this useful technique.

The predominance of pulse oximetry as the major commercial utilisation of PPG signals dictates that the bulk of existing attempts to address the problem of artefact are to be found within this context. These methods, which attempt to reduce the artefact or its effect on instrument outputs, have been examined for suitability of application to PPG studies in general and in terms of their inferred or reported success. It has been demonstrated that none of these techniques is ideal for fundamental PPG studies. It is therefore suggested that progress in understanding the origin and interpretation of PPG signals may be greatly improved by formulating a realistic physical model incorporating the major classes of artefact. In this way, desired signal components may be separated from artefact by using information obtained from the physical model to equalise the distortion. Because this approach removes the necessity to invoke *a-priori* assumptions about the characteristics of the PPG signal or artefact, it has the potential to produce truly general methods that could improve the accuracy and applicability of all PPG technologies.

2. MODELLING

Development of a photoplethysmographic model can be approached from a number of theoretical perspectives, necessitating a general description of signal artefact that may be applied in a multitude of contexts. The significance of this description for contemporary technologies may be explored by specific interpretations of the model, placing its insights within the framework of accepted mathematical foundations. One advantage of this approach is that physical significance can be attributed to different transduction techniques in a general manner, in this case a non-linear intensity measurement will be shown to aid interpretation. The consequences of non-linear transduction for the modelled form of artefacts can greatly simplify a description of the observed form of signal corruption, leading to simplifications in deterministic artefact reduction.

2.1 A HEURISTIC PPG MODEL

The primary physical assumption implicit in PPG studies is that changes in measured light intensity transmitted through or reflected from a vascular bed are induced by corresponding changes in illuminated blood volume. The fairly uniform distribution of blood within the peripheral vasculature suggests that we may model the effect of those blood volume changes as changes in the homogenous and macroscopic optical properties of the tissue. The tissue may be characterised according to the dictates of a suitable physical model for the optical interactions, whose exact form can be made dependent on the purpose of the study. Observed pulsations may then be attributed with physical significance by interpretation of the specific model. Observation, however, leads us to suppose that there will be certain aspects of the physical behaviour that can be described independently of the chosen model. It is therefore both possible and desirable to formulate a heuristic model based upon this behaviour, the physical interpretation of which can be made context dependent. Characteristic behaviour may then be quantitatively described in a general manner that is not constrained by the assumptions or applicability of a single physical model.

2.1.1 Definition of the model

Light is transmitted through or reflected from the anatomy of the patient, resulting in a measured intensity that depends upon the wavelength and intensity of the source, as well as any optical interaction with the subject. The primary assumption of this model is that the received intensity can be separated into components originating from different optical paths, therefore identifying and specifying distinct optical effects. We examine the case when a number of light sources can be used in conjunction with a single receiver to generate one or more PPG signals. The received light intensity as a function of time, t, and wavelength, λ, is modelled by

$$I(t,\lambda) = \sum_{j=1}^{n} I_j(t)\left[\alpha_j(t) + \beta_j(t,\lambda) + \gamma_j(t,\lambda)\right], \qquad \textbf{[2.1]}$$

where j labels the light source of intensity I_j and $\alpha_j, \beta_j, \gamma_j$ are coupling coefficients that depend on the geometric, temporal and spectral properties of the source / receiver positioning, artefact, tissue dynamics and the optical properties of both skin, blood and tissue. The coefficients α_j are interpreted as direct coupling between source and receiver and those labelled β_j, γ_j correspond to light coupled via non-pulsatile and pulsatile tissue respectively. The coefficients β_j encompass the coupling of light from any anatomical component that is not associated with dynamic arterial blood volume. In many instruments the light sources emit fixed power levels but here for the sake of generality they are allowed to vary. It is assumed here that the effects of ambient light have been successfully eliminated by an electronic multiplexing technique, an assumption that we will modify later. We now attribute the dynamic portion of the pulsatile tissue coupling coefficient, γ_j, to an underlying pulsatile signal $p_j(t)$,

$$\gamma_j(t,\lambda) = \gamma_j(\lambda) p_j(t). \qquad \textbf{[2.2]}$$

We can assume that the pulsatile signal represents a dynamic change in arterial blood volume and / or macroscopic optical properties of the tissue, which can ultimately be related back to a dynamic change in optical path length. It is this optical path length change that causes the dynamic modulation of the coupling coefficient expressed in equation [2.2], and therefore the observed intensity fluctuations in equation [2.1]. Specific interpretation of the detailed origin of this pulsatile signal in terms of the tissue properties or optical path length cannot be further investigated without recourse to a specific physical model. We will later explore the consequences of the choice of physical model on the interpretation of $p_j(t)$.

2.1.2 Applicability

The model developed in this section is underpinned by two fundamental assumptions:

1. The received intensity may be separated into components that can be attributed to the passage of light through various optical paths, with the optical coupling efficiency being free to change between distinct paths. There is no restriction on the physical model, which may vary between paths.

2. The arterial blood pulsations may be modelled as a change in optical path length that ultimately modulates the coupling of light from pulsatile tissue.

Although these assumptions are conceptually similar to those employed in conventional interpretation of PPG signals by technologies such as pulse oximetry, the generality is much improved by the heuristic nature of the model. Within the context of these assumptions, this heuristic model may be used to describe both venous and arterial PPG signals. The model does not dictate physical interpretation of the signals, nor does it attempt to explain the complex optical interactions.

2.2 INTERPRETATION OF THE MODEL

The model of equation [2.1] is simply a convenient method of describing the observed intensity and its fluctuations. The coupling coefficients identify different optical paths with different optical coupling efficiencies, and therefore separate the received intensity into optical-path dependent components. In order to attribute physical significance to these coefficients, it is necessary to invoke an analytic optical model that relates the observation of intensity fluctuations to the origin of the pulsations. There is no inherent reason why the model need be identical between the optical paths; indeed this generalisation will be used later in a new formulation for pulse oximetry. Initially, we will consider interpretation by homogenous models that describe the interaction of light with tissue identically for all optical paths. Two simple but useful examples will be presented, the first based upon the Beer-Lambert

law in order to aid understanding, and the second an example situation modelled by diffusion theory.

2.2.1 Choice of physical model

A mathematical description of the propagation and absorption of light in tissue can be constructed in two radically different ways: scattering theory and transport theory. Scattering theory is a fundamental approach that is based upon solution of Maxwell's equations for electromagnetic wave propagation. Although scattering theory can be used to obtain analytic solutions for simple geometric configurations, application to complex geometry or heterogeneous media can be mathematically complex. In contrast, transport theory heuristically models the transport of optical power through a general turbid medium, which is characterised with macroscopic optical properties. This technique is preferred by most authors because of its relative simplicity and extensive agreement with experimental evidence[77].

The Beer-Lambert law, which has already been discussed (section 1.3.2), is often used when a simple measure of optical density is required and can be shown to be a simplification of transport theory. The simplifying assumption (sometimes referred to as the *single scattering* approximation) is that there is no contribution to the forward direction of propagation from optical power within the medium. Although mathematically straightforward, the physical applicability of the Beer-Lambert law to highly scattering biological tissue is questionable[33,34].

The diffusion approximation to the transport equation treats optical propagation in terms of the diffusion laws within a homogenous medium[80] and is commonly used in the analysis of pulse oximetry[13,14,33,86]. The diffusion equation has shown good agreement with experimental results when the ratio of scattering to absorption is large or the scattering anisotropy of the medium is small[79]; the highly scattering condition being largely obeyed for the majority of biological tissue[83]. The diffusion approximation is more mathematically tractable than the underlying transport equation, enabling analytic solutions in a wider range of situations.

Monte-Carlo simulations offer a numerical modelling scheme that has shown good correlation with practical results in situations where an analytic solution is either unnecessary or impractical[78]. Numerical modelling of this type can be conceptually applied to either scattering or transport theory, and can also be applied to complex geometric configurations and heterogeneous media.

Since we are concerned primarily by the need to attribute physical significance to the heuristic parameters of our model, a numerical simulation may be of little use. We are therefore left with the Beer-Lambert law and diffusion theory as practical choices for an analytic description. Both of these approaches have been applied previously to various aspects of pulse oximetry (fundamental PPG studies usually employ numerical techniques).

2.2.2 Beer-Lambert interpretation

The Beer-Lambert law is used here to describe the case of transmission through tissue, as is common in pulse oximetry. The resulting measured transmitted intensity can be expressed as $I_0 T(z,\mu)$, where I_0 is the incident intensity and $T(z,\mu)$ is the *transmittance* of the tissue, which is dependent on the physical path length, z, and the effective absorbance, μ. The transmittance is expressed by reference to the form of the Beer-Lambert law (equation [1.2]),

$$T(z,\mu) = \exp(-z\mu). \qquad [2.3]$$

2.2.2a Separation of signal components

The transmitted intensity will consist of a static component and a dynamic component that is primarily due to the arterial pulsations and therefore, in this interpretation, dynamics in the homogenous and macroscopic optical properties of the tissue. We can therefore relate these components back to our heuristic model and write the total transmittance, through both pulsatile and non-pulsatile tissue, in terms of the previously defined coupling coefficients,

$$\beta + \gamma p(t) = T\big(z + \Delta z(t), \mu + \Delta\mu(t)\big), \qquad \textbf{[2.4]}$$

where we have assumed that there is no direct coupling of light between source and receiver. We have now attributed some physical significance to the pulsatile signal, $p(t)$, by modelling its effect as a dynamic change in optical path length due to either a dynamic change in physical path length or in the macroscopic effective abosrbance of the tissue. Because $\Delta z(t)$ and $\Delta\mu(t)$ are small by comparison to their static counterparts, we may linearise equation [2.4] with respect to these small changes using a first order Taylor expansion,

$$T\big(z + \Delta z(t), \mu + \Delta\mu(t)\big) \approx T(z,\mu) + \Delta z \frac{\partial T(z,\mu)}{\partial z} + \Delta\mu \frac{\partial T(z,\mu)}{\partial \mu}, \qquad \textbf{[2.5]}$$

which enables us to write

$$\beta + \gamma p(t) \approx \exp(-z\mu)\big[1 - \Delta z(t)\mu - \Delta\mu(t) z\big]. \qquad \textbf{[2.6]}$$

Since β represents the static part of the total transmittance, $T(z,\mu)$, we can rewrite equation [2.6] as a ratio of the dynamic to static components,

$$\frac{\gamma p(t)}{\beta} = -\Delta z(t)\mu - \Delta\mu(t) z\,. \qquad \textbf{[2.7]}$$

2.2.2b Interpretation

Equation [2.7] enables physical interpretation of the ratio of coupling coefficients, γ/β, in terms of the static parameters and the specific model for the origin of observed pulsations. For example, the common derivation for arterial oxygen saturation in pulse oximetry models the observed pulsations as changes in the physical path length through blood only (see section 1.3.2), implying that the total transmittance may be written as

$$T\left(\mu_{tissue}, r, \mu_{blood}, z+\Delta z(t)\right)=\exp-\left(\mu_{tissue} r+\mu_{blood}\left(z+\Delta z(t)\right)\right), \quad \textbf{[2.8]}$$

where r is the static path length through tissue and it is implicit that $p(t)=\Delta z(t)$.

In this case equation [2.7] reduces to

$$\frac{\gamma p(t)}{\beta}=-\Delta z(t)\mu_{blood}. \quad \textbf{[2.9]}$$

When using the physical model and assumptions employed for the pulse oximetry derivation, equation [2.9] tells us that the ratio of coupling coefficients is now equal to the effective absorbance of the blood only. The ratio of ratios used to calculate oxygen saturation is defined as the ratio of the absorbance of the illuminated blood at two wavelengths. It can therefore be seen that a ratio of two instances of equation [2.9] could be used to rewrite the oxygen saturation in terms of our heuristic coupling coefficients. This interpretation will be utilised later to analyse the effects of modelled artefacts on a conventional formulation of pulse oximetry, from examination of the behaviour of the heuristic model.

This useful example serves to illustrate how physical significance can be attributed to the components of our heuristic model in a specific context. Observation of the behaviour of these heuristic components may then be extrapolated back to their physical counterparts, providing insight that would otherwise be difficult to obtain. The use of the Beer-Lambert law has served to enable us to place the model within the context of standard pulse oximetry derivations, whilst maintaining a high degree of simplicity.

2.2.3 Diffusion theory interpretation

The diffusion approximation to the Boltzmann transport equation treats optical scattering in terms of the diffusion laws within a homogenous medium[80], with

diffusion, D (m), scattering, μ_s (m^{-1}), and absorption, μ_a (m^{-1}), coefficients defining the optical properties of that medium. Analytic description of optical interactions with highly scattering tissue is commonly achieved with diffusion theory[83], and its application to the analysis of pulse oximetry is widespread[13,14,33,86]. Experiments have verified that diffusion theory is applicable to scattering by human blood[81] for transmission distances of greater than 1mm with a haematocrit level (percent by volume of packed red blood cells in whole blood) greater than 0.05; conditions which should both be satisfied in the context of PPG studies. In addition diffusion theory has been shown to be a good approximation for the optics of human skin[82]. Diffusion theory can equally be applied to transmission through or reflection from tissue, however the subsequent discussion will be limited to the transmission mode case for the sake of simplicity.

In steady state conditions, the spatial distribution of the scattered photon density $\phi(\mathbf{r})$, or energy fluence rate (Wm^{-2}), satisfies the diffusion equation

$$D\nabla^2\phi(\mathbf{r}) - \mu_a\phi(\mathbf{r}) = -S(\mathbf{r}), \qquad \textbf{[2.10]}$$

$$D = \frac{1}{3}\left[\mu_a + \mu_s(1-g)\right]^{-1}. \qquad \textbf{[2.11]}$$

Here g is the mean cosine of the scattering angle of a single scattering event, or the anisotropy of the medium. $S(\mathbf{r})$ is a source term, usually generated by the scattering of collimated irradiation, i.e. that fraction of an incident beam that has been neither scattered nor absorbed. Consideration of the transport equation indicates that this collimated source term should decrease exponentially with distance into the scattering medium[83]. For diffuse illumination, the source term is usually set to zero with the diffuse incidence being introduced in the boundary conditions.

2.2.3a Semi-infinite slab approximation

For the majority of biological tissue, scattering dominates the transport process, with values for the diffusion constant, D, being less than 1mm[83]. In this case the unscattered or collimated term tends to zero after only a few diffusion lengths and can then practically be ignored[84]. Here a semi-infinite slab of homogenous medium is considered with a point source placed just inside the medium at a depth where the first scattering events would occur for collimated irradiation[84]. This localisation of the first interactions gives rise to equation [2.12], which uses the Dirac delta function to represent the scattered point source.

$$S(\mathbf{r}) = P_0 \delta(\mathbf{r} - \mathbf{r}'), \qquad \textbf{[2.12]}$$

where $\mathbf{r}'$ locates the point source and P_0 represents the source power. Substitution of this source term into equation [2.10] enables a Green's function solution[85] for the photon density,

$$\phi(\mathbf{r}) = \frac{P_0}{4\pi D} \frac{\exp\left(-\sqrt{\frac{\mu_a}{D}} R\right)}{R}, \qquad \textbf{[2.13]}$$

where $R = |\mathbf{r} - \mathbf{r}'|$ is the scalar distance between $\mathbf{r}$ and the point source. Note that if the origin of the co-ordinate system is defined to be coincident with the point source, then $R = r = |\mathbf{r}|$.

2.2.3b Detection

An isotropic detector placed inside the medium will have an output proportional to $\phi(\mathbf{r})$. If we wish to describe the behaviour of transmitted light through a medium, however, we must consider the scattered photon flux density in the transmission direction[84]. For example, a PiN diode receiver placed at the skin surface will

primarily be sensitive to photon flux in the direction of the detector surface normal, i.e. the normal to the skin. The directional flux term is obtained from Ficks law[80],

$$J(\mathbf{r}) = -D\nabla\phi(\mathbf{r}), \qquad \textbf{[2.14]}$$

and defining co-ordinates such that the detector surface is normal to the positive z direction, consideration of equation [2.14] with respect to this dimension only gives

$$J_z(\mathbf{r}) = \frac{P_0}{4\pi} z \exp\left(-\frac{R}{\delta_d}\right)\left[\frac{1}{\delta_d R^2} + \frac{1}{R^3}\right], \qquad \textbf{[2.15]}$$

where the substitution $\delta_d = \sqrt{D/\mu_a}$ has been made for convenience. We may integrate the flux of equation [2.15] over the area, A, of a detector of quantum efficiency η to yield the detected power,

$$P(\mathbf{r}) = P_0\eta\frac{z}{4\pi}\iint_A \exp\left(-\frac{R}{\delta_d}\right)\left[\frac{1}{\delta_d R^2} + \frac{1}{R^3}\right]dxdy. \qquad \textbf{[2.16]}$$

Equation [2.16] may be simplified by placing the detector such that it is directly above the source, $x - x' = y - y' = 0$. If we further assume that the detector surface area, A, is small by comparison to the source detector separation, R, then we can approximate R by z. The measured power therefore becomes

$$P(z) \approx \frac{P_0\eta A}{4\pi z}\exp\left(-\frac{z}{\delta_d}\right)\left[\frac{1}{\delta_d} + \frac{1}{z}\right]. \qquad \textbf{[2.17]}$$

It can be seen from equation [2.17] that the optical path length is a function of z and the tissue optical characteristics, δ_d and μ_s. Some authors[81] refer to δ_d as the penetration depth for unscattered photons.

2.2.3c Interpretation

We now observe that a measured transmitted intensity, $z^2 P(z, \delta_d, \mu_s)/A$ (Wsr^{-1}), has a static and a dynamic portion. We can therefore say that the parameters defining the optical path length also have a static and a dynamic portion. Referring to equations [2.1] and [2.2], we can equate the diffuse transmittance to the heuristic coupling coefficients,

$$\beta + \gamma p(t) = \frac{1}{P_0} P\big(z + \Delta z(t), \delta_d + \Delta\delta_d(t), \mu_s + \Delta\mu_s(t)\big), \qquad \textbf{[2.18]}$$

where the dynamic changes, $\Delta z(t)$, $\Delta\delta_d(t)$ and $\Delta\mu_s(t)$, are small by comparison to their static counterparts and it has been assumed that there is no direct coupling of light between source and receiver. Equation [2.17] may therefore be linearised with respect to these small changes,

$$\begin{aligned}\beta + \gamma p(t) \approx \frac{\eta A}{4\pi z}\exp\left(-\frac{z}{\delta_d}\right)\left[\frac{1}{\delta_d}+\frac{1}{z}\right] \\ \left[1-\frac{\Delta z(t)}{z}\left(2+\frac{z^2}{\delta_d(z+\delta_d)}\right)-\frac{\Delta\delta_d(t)}{z}\left(1+\frac{\delta_d}{z}\right)+\frac{\Delta\mu_s(t)}{\mu_s}\right].\end{aligned} \qquad \textbf{[2.19]}$$

Since β represents the static component, $P(z, \delta_d, \mu_s)/P_0$, equation [2.19] may be rewritten in terms of the ratio of the dynamic to static components,

$$\frac{\gamma p(t)}{\beta} = -\frac{\Delta z(t)}{z}\left(2+\frac{z^2}{\delta_d(z+\delta_d)}\right)-\frac{\Delta\delta_d(t)}{z}\left(1+\frac{\delta_d}{z}\right)+\frac{\Delta\mu_s(t)}{\mu_s}. \qquad \textbf{[2.20]}$$

2.2.3d Simplified interpretation

In the derivation of equation [2.20], we have invoked a homogeneous model to describe the essentially heterogeneous physical reality of the human vasculature. This approximation is necessary because of the unpredictable nature of the

heterogeneity. In conversion from the heterogeneous reality to the homogeneous model, we may choose to ascribe the physiological dynamics to any combination of the parameters described in equation [2.20]. In a tissue medium where scattering dominates the transport process, the dynamic term in μ_s is unlikely to contribute significantly to the overall dynamics. It is evident that z is much larger than δ_d, indeed this must be true for the point source approximation to be valid, which implies that a dynamic change in z will dominate the ratio $\gamma p(t)/\beta$. It is therefore reasonable to model the underlying heterogeneous dynamics as dynamic changes in the physical path length, z, of the homogeneous model. We can thus model $p(t)$ as a change in physical path length, $\Delta z(t)$, which enables us to write

$$\frac{\gamma p(t)}{\beta} \approx -\frac{\Delta z(t)}{z}\left(2+\frac{z^2}{\delta_d\left(z+\delta_d\right)}\right). \qquad \textbf{[2.21]}$$

We can now assign a physical meaning, in terms of the static tissue parameters of our model, to the static ratio γ/β,

$$\frac{\gamma}{\beta} \approx -\frac{1}{z}\left(2+\frac{z^2}{\delta_d\left(z+\delta_d\right)}\right). \qquad \textbf{[2.22]}$$

In a spectrophotometric interpretation of equation [2.22] for pulse oximetry or another applied technology, δ_d characterises the physiological component being investigated with wavelength dependency.

2.2.3e Conclusions

This analysis has been performed in order to illustrate one example of the physical significance of the heuristic model. Other interpretations are possible, but it is the linearisation of the relatively small dynamic portion of the model that has the most significance in the context of our heuristic model. Just as in the simplistic Beer-Lambert law example, we have related the parameters of our choice of physical

model to the parameters of the heuristic model. This enables us to physically interpret the observed intensity dynamics in terms of the model for the origin of the pulsations and characteristics of the tissue. Observed behaviour of the heuristic model can also be used to formulate some plausible conclusion as to the behaviour of either physical model under similar conditions.

2.3 NON-LINEAR RESPONSE CHARACTERISTIC

Whilst the received intensity of equation [2.1] can provide useful information about light coupled from various optical paths, it is now proposed that a non-linear measure of this received intensity be used. A non-linear measurement enables complex functional dependencies to be expressed in terms of linear equations. For example, a logarithmic transformation of the dependent variable can be used to rewrite a power law in a linear fashion, or re-express a proportional term in an additive way. The natural logarithm is the obvious choice for producing such a non-linearity, because it is easily implemented electronically. In addition, PiN diode receivers can be biased in such a way that they automatically give a logarithmic measure of the incident intensity. Whilst other non-linear transformations are possible, their practical implementation often relies upon a base of logarithmic circuitry. In addition, the logarithmic transformation has a useful property when applied to signals containing a large static component and a small dynamic component, as is the case for PPG signals.

Consider a signal, s, which is a linear representation of a measured intensity. This signal is the sum of a large static component, s_{DC}, and a small dynamic component, s_{AC}. We can express the logarithm of this signal as

$$\log(s) = \log(s_{DC}) + \log\left(1 + \frac{s_{AC}}{s_{DC}}\right), \qquad \textbf{[2.23]}$$

and by further noting that the ratio s_{AC}/s_{DC} is small because $s_{DC} \rangle\rangle s_{AC}$, we can use the relation $\log(1+x) \approx x$ as $x \to 0$, to obtain

$$\log(s) \approx \log(s_{DC}) + \frac{s_{AC}}{s_{DC}}. \qquad \textbf{[2.24]}$$

Therefore the logarithmic transformation of such a signal has the effect of re-normalising the dynamic component with the static signal size. This property of logarithmic amplifiers has been used electronically to implement automatic gain control systems. We introduce it here because practical PPG signals have very small dynamic components in respect of their static magnitudes, and therefore fit this approximation very well.

2.3.1 Logarithmic characteristic

We shall now consider the effect of a non-linear receiver characteristic, such as the logarithmic response obtained from an open-circuit PiN diode, on our heuristic model. Alternatively, a linear receiver could be used in conjunction with a subsequent logarithmic amplifier. The intensity falling on the receiver is assumed to have the form given in equation [2.1]. We may then define a dynamic signal, $v(t,\lambda)$, to represent a logarithmic measure of the received intensity,

$$v(t,\lambda) = v_0 \ln\left(1 + i_{ph}(t,\lambda)Z\right), \qquad \textbf{[2.25]}$$

where the photocurrent, i_{ph}, has wavelength dependency through the optical transducer responsivity, Z is a constant characteristic of the receiver, and v_0 includes any electronic gain applied. The photocurrent is linearly proportional to the ambient-subtracted light intensity incident on the receiver, with the additive constant modelling the possibility of a receiver dark current or offset. Substitution of equation [2.1] into equation [2.25] yields

$$v(t,\lambda) = v_0 \ln\left(1 + ZR_j(\lambda) I_j\left(\alpha_j + \beta_j(\lambda) + \gamma_j(\lambda) p_j(t)\right)\right), \qquad \textbf{[2.26]}$$

where R_j is introduced as the wavelength dependent responsivity to a constant light source of intensity I_j. Equation [2.26] expresses the possibility of a non-linear measure of the intensity, whilst maintaining a high degree of generality. This formulation is valid whether the practical implementation uses non-linear transduction or logarithmic signal processing following linear transduction. This will prove useful later on, when we can consider implementation issues and possible simplifications without requiring a re-formulation of the model.

2.3.2 Self calibrating action

Since the pulsatile component is observed to be small by comparison to the static component, equation [2.26] may be approximated to first order in the size of the pulsatile signal, by first rewriting in the form

$$v(t,\lambda) = v_0 \ln\left[\left(1 + ZR_j(\lambda) I_j\left(\alpha_j + \beta_j(\lambda)\right)\right)\left(1 + \frac{ZR_j(\lambda) I_j \gamma_j(\lambda) p_j(t)}{1 + ZR_j(\lambda) I_j\left(\alpha_j + \beta_j(\lambda)\right)}\right)\right], \qquad \textbf{[2.27]}$$

noting that the receiver offset is much less than all other signal sources, $ZR_j(\lambda) I_j\left(\alpha_j + \beta_j(\lambda)\right) \gg 1$, and that the dynamic coupled term is much smaller than the static term, $\left(\alpha_j + \beta_j(\lambda)\right) \gg \gamma_j(\lambda) p_j(t)$, to give

$$v(t,\lambda) \approx v_0 \ln\left[1 + ZR_j(\lambda) I_j\left(\alpha_j + \beta_j(\lambda)\right)\right] + v_0\left(\frac{\gamma_j(\lambda) p_j(t)}{\alpha_j + \beta_j(\lambda)}\right). \qquad \textbf{[2.28]}$$

It may be seen from equation [2.28] that the received logarithmic signal consists of a static and a dynamic component. Whilst the static component is written as the logarithm of the linear static component, the simplifications that have arisen from consideration of the relative sizes of terms have enabled us to write the dynamic component in a linear manner.

2.3.2a Implications

The dynamics of equation [2.28] have been modified by the non-linearity, with respect to a linear measurement, in a number of ways:

1. The dynamics are now independent of the source intensity, receiver sensitivity and logarithmic transformation characteristics.

2. The received pulsations have been re-normalised with the non-pulsatile tissue coupling co-efficient, β. This implies that, in the case of negligible coupling between source and receiver, the observed dynamics may be directly related to equation [2.9], [2.21] or another relation derived from a different model.

3. A scaling factor, v_0, has been applied to the dynamics.

Thus we can say that the non-linear transformation has a self-calibrating action, which normalises the received pulsations with optical information in such a way that we can now interpret the observed fluctuations with physical significance. The independence of source intensity and receiver characteristics enable quantitative comparison of pulsations, in a similar manner to the technique of controlling the source intensity to maintain a pre-determined constant average received value[15]. In addition, interpretation of the heuristic model has demonstrated the importance of the static ratio γ/β for tissue characterisation according to a specific physical model. This implies that technologies based upon interpretation of PPG signals could be simplified by this self-calibration.

2.3.3 Interpretation of resulting signals

This self calibrating action can be demonstrated by consideration of the dynamic portion of equation [2.28] under the conditions of zero direct coupling and the homogeneous physical path length model of equation [2.21],

$$v(t,\lambda)_{AC} \approx v_0 \frac{\Delta z_j(t)}{z_j}\left(2+\frac{z_j^{\,2}}{\delta_d(\lambda)(z_j+\delta_d(\lambda))}\right), \qquad \textbf{[2.29]}$$

the Beer-Lambert formulation of equation [2.7] ,

$$v(t,\lambda)_{AC} \approx v_0\left(-\Delta z_j(t)\mu_j(\lambda)-\Delta\mu_j(t,\lambda)z_j\right), \qquad \textbf{[2.30]}$$

or a similar relation derived from another physical model. It has now become possible to directly relate the observed transformed dynamics to the modelled source of pulsations. A formulation performed without the non-linearity would not yield the same independence of source intensity or detector characteristics and would not be linear in the static tissue characteristics. In addition, this self-calibrating action, by effectively incorporating optical information about the static component into the dynamic component, implies that many types of analysis need only consider the dynamic component of a transformed signal. For example, physiological characterisation using either physical model shown above (equations [2.29] and [2.30]) can be performed by consideration of a pair of transformed dynamics at different wavelengths. This self-calibrated signal will also prove useful when interpreting the effects of any additional signal dynamics, such as the effects of motion artefact.

2.4 MODELLING THE EFFECTS OF MOTION ARTEFACT

One of the consequences of utilising a heuristic model to describe the observed intensity is that it enables us to model artefact in a heuristic manner. This is useful because artefact may be extremely complex in origin and a detailed physical model may not be practical or generally applicable. Observation of various artefacts of differing physical origin can be used to build a picture of how artefact manifests itself within the context of the heuristic model. Whilst it is straightforward to model residual ambient artefact as an additional intensity, we must decide on a functional

form for a model of motion artefact. A number of possible motion artefact models have already been introduced, which can be said to fall into three categories:

- Additive – Artefact is assumed to be an additional intensity component that is independent of the optical path and does not modify the received intensity due to the underlying physiological dynamics.

- Signal characterisation – Artefact is assumed to be identifiable from the desired dynamics by characterisation of the spectral or shape-based temporal characteristics of the resulting signals.

- Multiplicative – Artefact is assumed to modulate all components of the received intensity in an equal manner.

Of the possible choices for a heuristic artefact model, the additive and multiplicative scenarios offer the greatest scope for generality. An additive artefact model can easily be incorporated into our heuristic model with an additional term, whereas any multiplicative effects may be described as modulation of the coupling coefficients from various optical paths. Such a model will therefore be independent of the specific spectral and temporal characteristics of either the physiological dynamics or artefact.

2.4.1 Model determination

It is possible to experimentally determine which of the two postulated models would be the most physically appropriate, by performing artefact-inducing experiments. In a stable environment, with no artefact of any kind, the magnitudes of the static (dc) and dynamic (ac_{pp}) portions of the received intensity from a single-channel PPG system are recorded. An artificial artefact is then induced, using a predetermined experimental protocol. During the period when the artefact is induced the static (dc') and dynamic (ac'_{pp}) intensity magnitudes are again recorded. In the general case, we can model the effect of motion artefact as both a multiplicative factor, $(1+m)$, and an

additive term, a. Assuming that the artefact-inducing conditions remain static, we can relate the measured signal components before and after induction of the artefact,

$$ac'_{pp} = ac_{pp}(1+m), \qquad \textbf{[2.31]}$$

$$dc' = dc(1+m)+a\,. \qquad \textbf{[2.32]}$$

We can then define a normalised dynamic magnitude, $ac'_{normalised} = ac'_{pp}/dc'$, which, in the absence of an additive term, would be independent of any multiplicative factor that affects both signal components in an identical manner. In the case of a finite additive term, the normalised dynamic magnitude during artefact induction can be expressed in terms of the signal components before induction and our heuristic artefact terms,

$$\frac{1}{ac'_{normalised}} = \frac{dc}{ac_{pp}} + \frac{a}{ac_{pp}(1+m)}. \qquad \textbf{[2.33]}$$

The ratio of normalised dynamic magnitudes before and after induction of the artefact is then

$$\frac{ac_{normalised}}{ac'_{normalised}} = 1 + \frac{a}{dc(1+m)}. \qquad \textbf{[2.34]}$$

This ratio will therefore be unity if artefacts manifest themselves as multiplicative factors that affect all signal components in an identical manner, with the additive artefact term set to zero. Other numerical values of this ratio would suggest the presence of an additive artefact effect.

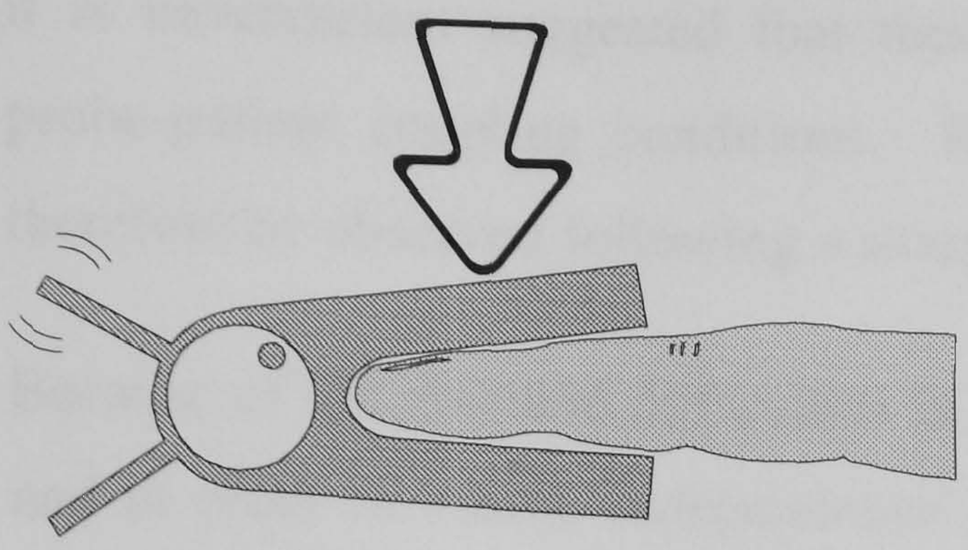
2.4.2 Physical investigation

In order to investigate and physically justify a specific model for motion artefact, it is necessary to obtain experimental evidence from realistic artefact situations. For this

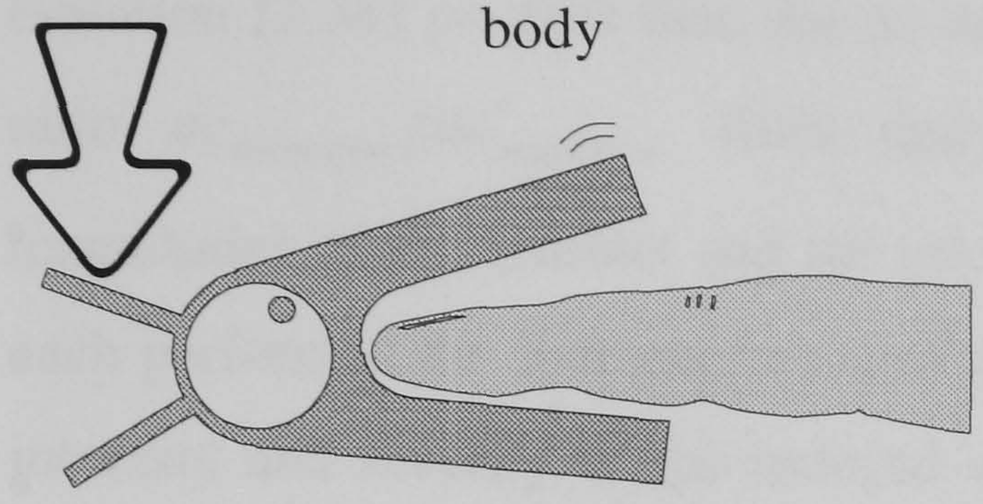
purpose a pulse oximetry finger probe was used as a source of transmission mode PPG signals. The probe was driven using specially designed electronics to provide a fundamental source of signals that had not been processed in any way. Static and dynamic signal components were then separated using electronic filtering and fixed gain was applied to the small dynamic signal to improve the signal-to-noise ratio. Artefacts could then be produced under laboratory conditions and their effect on the magnitude of the static and dynamic components evaluated.

2.4.2a Experimental methodology

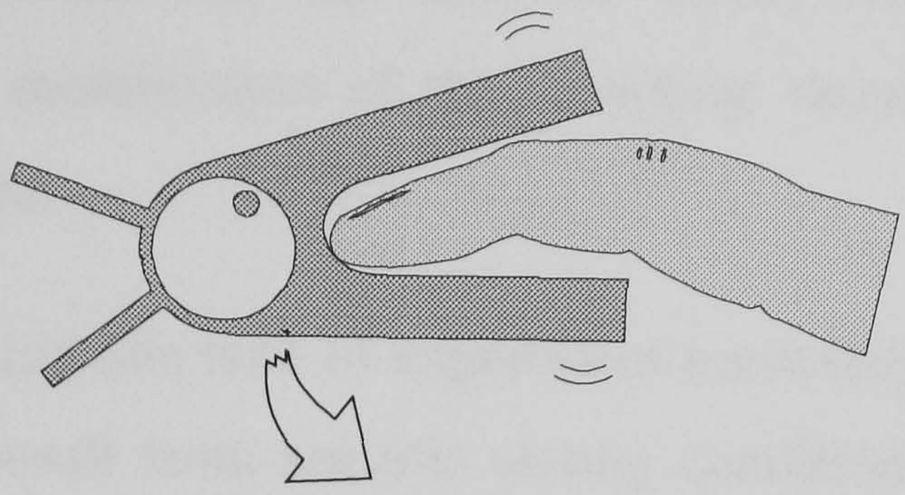
Since it is difficult to analyse the detailed temporal characteristics of high-speed artefact, a number of steady-state tests were performed that were designed to enable objective isolation of the signal components (see figure 2.1). In the first test, light pressure was applied to the body of the probe, simulating a common effect in clinical practice. Care was taken not to apply enough pressure to compress the tissue, which would result in unpredictable changes in the measured physiological dynamics. Instead, the effect of such pressure is to marginally alter the physical coupling between the probe opto-electronics and the patient's skin. Immediately following the application of such pressure there will be a transient effect that is partially due to the artefact itself and partially due to the settling time of the electronic filters and associated circuitry. In the test the pressure was therefore maintained until clean arterial signals were again obtained, enabling a quantitative comparison of the magnitudes of the static and dynamic components before and after application of the pressure. Alternatively, performing slight movements of the finger within the probe could be used to simulate internal pressure changes. Although the artefact produced with this method is much slower than would be expected in a realistic situation, the analogy is valid because the model does not depend upon the temporal or spectral characteristics of the various components.



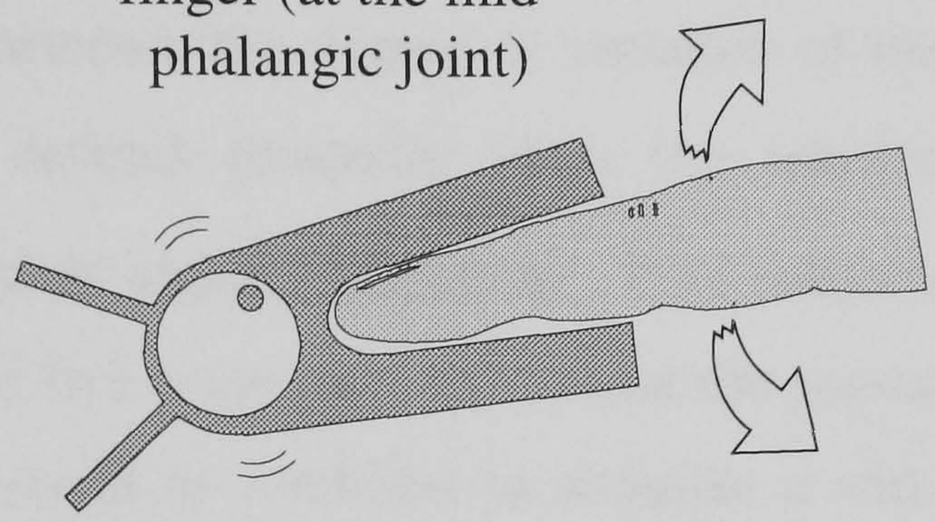
Test 1 – Application of pressure to the probe body



Test 2 – Application of pressure to the probe clip



Test 3 – Bending the finger (at the mid-phalangic joint)



Test 4 – Waving movements of the hand

Figure 2.1 – Artefact inducing experimental protocols

The second test performed extended the idea of modifying the probe coupling, by producing artificial artefacts that could only be due to small movements between the subject and the probe opto-electronics. Application of light pressure to the probe clip was used to produce such artefact, by reducing pressure on the subject and moving the opto-electronics fractionally away from the skin. Again, conditions of stable measurement were obtained before comparison of the resultant signals.

For the third test, production of another type of artefact was achieved by bending the finger to which the probe was attached. This is likely to be a common artefact in clinical environments, where dynamic movement of the hand causes a varying degree of bending of the finger, which in turn produces slight movement of the finger within the body of the probe.

In the final test, sharp movements of the hand simulated another common artefact encountered in a clinical environment. Whilst the transient effects of this type of movement may differ from the effects observed after the patient has stopped moving,

it is nevertheless suggested that these movements will still alter the steady-state probe-patient coupling conditions. Residual modification of this coupling should therefore be observed following a sharp transient.

Because of the practical difficulties in performing this type of experiment accurately, and in order to ensure independence of any result from specific testing conditions, one hundred trials were performed and a probability distribution was constructed. Equation [2.34] predicts that, for an additive artefact, the degree of variation of the ratio $ac_{normalised}/ac'_{normalised}$ from unity will depend primarily upon the relative magnitudes of the artefact and the uncorrupted dc signal component. Five subjects each performed the four artefact tests a total of five times per test, so that the source intensity and severity of the induced artefact could be modified to simulate a wide range of values for the quantity $a/dc(1+m)$ for each test. Whilst the source was electronically varied to produce random intensities (and therefore random dc values), the severity of induced artefact had to be judged manually. We would therefore expect an approximately uniform probability distribution for the ratio if an additive artefact model were physically appropriate. A distribution centred about unity and having a low standard deviation would suggest a multiplicative model.

2.4.2b Results

Figure 2.2 shows a probability distribution function for the ratio $ac_{normalised}/ac'_{normalised}$ recorded from these tests. The mean value for this distribution is 1.06, with a standard deviation of 0.17. Although it is difficult to obtain enough variation in the size of artefact produced to give objective results, the strong grouping towards unity indicates that a multiplicative artefact model would be more physically appropriate than an additive model.

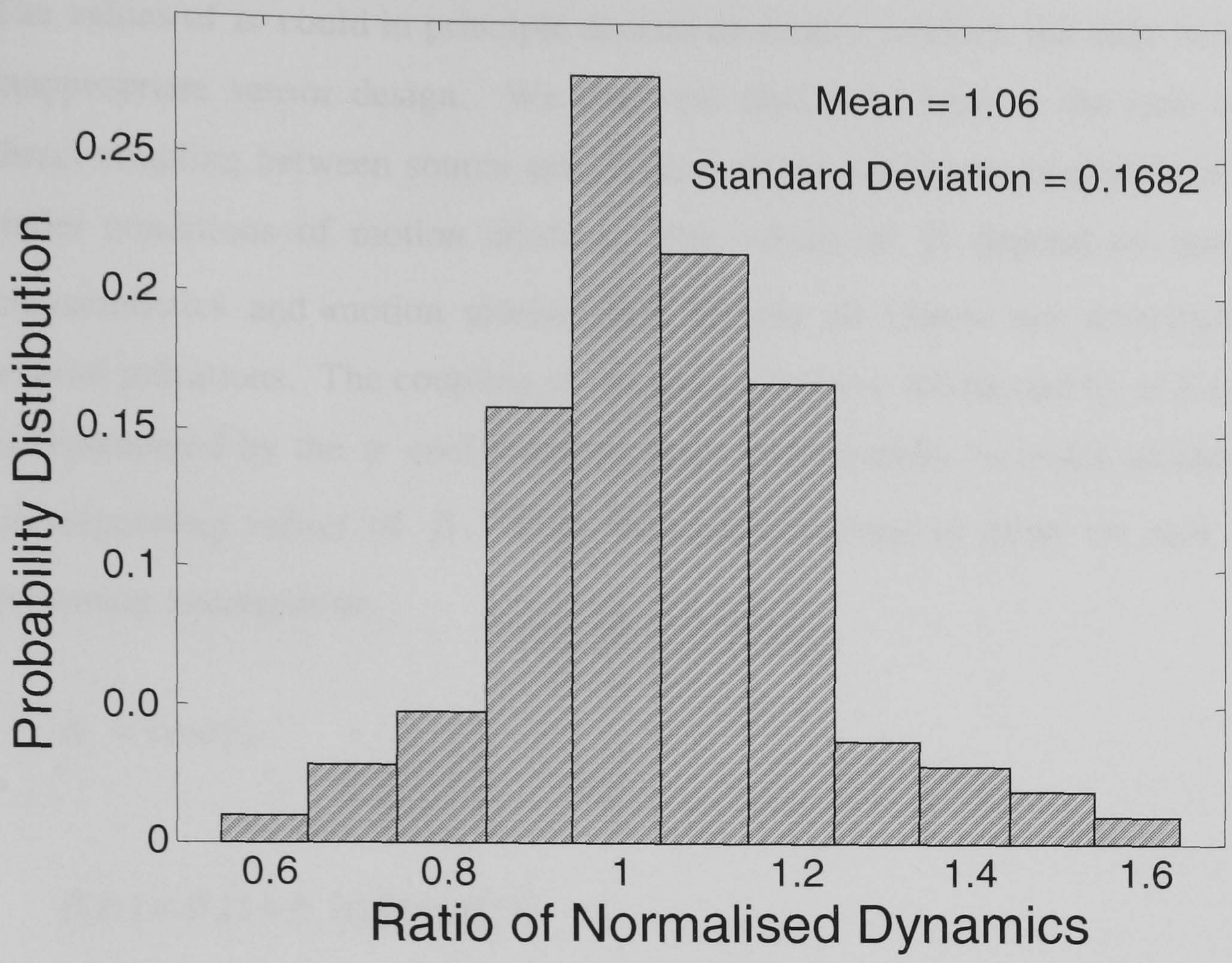


Figure 2.2 – Probability distribution of ratio of normalised dynamics

Whilst these measurements do not exactly define the functional form of motion artefact, they do serve to indicate that a multiplicative model is more appropriate than the commonly used additive model. This is a physically plausible conclusion, because a multiplicative model acknowledges that the artefact will affect the received intensity from both static tissue and arterial pulsations. It is suggested that the predominant effect of motion artefact is to modify the optical coupling between probe and tissue, an effect we shall call *probe-coupling artefact*.

2.4.3 A heuristic artefact model

Since experimental observation indicates that a multiplicative artefact model is the most physically appropriate, we can now examine the nature of the coupling coefficients (α, β and γ) under conditions of motion artefact more closely. Modulation of these coefficients can introduce path-dependent multiplicative effects, which can then be attributed to one or more physical sources of artefact.

The values of α could in principle depend on motion artefact, but only in the case of inappropriate sensor design. We limit the discussion here to the case where any direct coupling between source and detector is assumed to be negligible or invariant under conditions of motion artefact. The values of β depend on tissue optical characteristics and motion artefact and include all effects not attributable to the arterial pulsations. The coupling of arterial pulsations, modulated by motion artefact, is represented by the γ coefficients, which will normally be much smaller than the corresponding values of β. With these observations in mind we now make the following assumptions,

$$\alpha_j = const\,, \qquad \textbf{[2.35]}$$

$$\beta_j(t) = \beta_j\big(1+b_j(t)\big)\big[1+m(t)\big]\,, \qquad \textbf{[2.36]}$$

$$\gamma_j(t) = \gamma_j\big[1+m(t)\big]p_j(t)\,, \qquad \textbf{[2.37]}$$

enabling the received intensity (equation [2.1]) to be written as

$$I(t,\lambda) = \sum_{j=1}^{n} I_j(t)\left[\alpha_j + \beta_j(\lambda)\big[1+m(t)\big]\left(1+b_j(t)+\frac{\gamma_j(\lambda)}{\beta_j(\lambda)}p_j(t)\right)\right]. \qquad \textbf{[2.38]}$$

The assumption [2.35] is valid for probes in which the opto-electronics have fixed relative positions such that there is no direct coupling of light. Equation [2.36] distinguishes motion artefact $m(t)$ due to changes in probe coupling from dynamics arising due to changes in non-pulsatile blood volume $b(t)$. This is important in distinguishing motion artefact in the context of pulse oximetry, in which $b(t)$ is considered to be artefact, from the context of PPG in which $b(t)$ is considered useful information (see section 1.4.1a). Assumption [2.37] is justified by criteria similar to those for assumption [2.36] with the addition of independence between the pulsation

dynamics and the artefact dynamics. The assumptions [2.35] to [2.37] model both probe-coupling motion artefact and dynamic venous blood volumes as linear factors that modulate the coupling coefficients, maintaining generality and being in agreement with our assumption of multiplicative artefact. The functionality of these factors is assumed to be constant, and does not depend upon the spectral characteristics of either the PPG signal or the artefact. We have therefore incorporated the possibility of multiplicative artefact into the framework of our heuristic model, whilst maintaining a high degree of generality.

2.4.3a Consequences of the model

There are a number of distinct elements of the model in equation [2.38] that can dominate the final signal and it is important to analyse the detailed ordering of terms. Taking the case when the direct coupling of light is negligible and neglecting dark current with respect to all other current sources, we can write the logarithmically transformed intensity (equation [2.28]) as

$$v_j(t,\lambda) = v_0 \ln\left(ZR_j I_j \beta_j\right) + v_0 \ln\left(1+m(t)\right) + v_0 \ln\left(1 + b_j(t) + \frac{\gamma_j(\lambda) p_j(t)}{\beta_j(\lambda)}\right). \qquad \textbf{[2.39]}$$

It can be seen that the assumptions [2.35], [2.36] and [2.37], along with the non-linear response, lead to a separation of probe coupling motion artefact into a single term that is independent of the light source intensity. This implies that the artefact term may be deterministically removed from the desired signal, therefore achieving motion artefact reduction or *equalisation*. The difference between a pair of instances of equation [2.39] at different wavelengths, for example, would be independent of the motion artefact term. The detailed implementation of deterministic artefact reduction will be considered in subsequent sections.

2.4.3b Consequences for pulse oximetry

It has previously been demonstrated that pulse oximeters determine the arterial oxygen saturation from consideration of two PPG signals at different wavelengths. A

quantity called the ratio of ratios can be calculated from the received intensities, which can be empirically calibrated against the invasively verified oxygen saturation (see section 1.3). There are two predominant methods of calculating the ratio of ratios, R, from two PPG intensities (the logarithmic and the linear algorithms), which equivalently give the absorbency ratio of blood at the two wavelengths,

$$R = \frac{\mu_{blood}(\lambda_1)}{\mu_{blood}(\lambda_2)} = \frac{\dfrac{d[\ln I(t,\lambda_1)]}{dt}}{\dfrac{d[\ln I(t,\lambda_2)]}{dt}} = \frac{I(t,\lambda_2)\dfrac{d[I(t,\lambda_1)]}{dt}}{I(t,\lambda_1)\dfrac{d[I(t,\lambda_2)]}{dt}}. \qquad \textbf{[2.40]}$$

Interpretation of the heuristic PPG model using a simple Beer-Lambert law has also been used to attribute physical significance to the re-normalised dynamics of a logarithmically transformed signal (equations [2.9] and [2.28]). In this case we can interpret the ratio of ratios in terms of the dynamics of a pair of transformed signals,

$$\frac{v(t,\lambda_1)_{AC}}{v(t,\lambda_2)_{AC}} \approx \frac{\gamma(\lambda_1)\beta(\lambda_2)}{\beta(\lambda_1)\gamma(\lambda_2)} = \frac{\mu_{blood}(\lambda_1)}{\mu_{blood}(\lambda_2)} = R. \qquad \textbf{[2.41]}$$

It is a simple matter to demonstrate that this technique is equivalent to the formulations of equation [2.40], by noting that

$$\frac{d[\ln I(t,\lambda)]}{dt} = \frac{d[v(t,\lambda)_{AC}]}{dt} \approx \frac{\dfrac{d[I(t,\lambda)]}{dt}}{I(t,\lambda)}, \qquad \textbf{[2.42]}$$

and observing that the derivative of an AC coupled signal can always be expressed as the difference between the instantaneous AC signal and the DC level, in other words the AC signal itself. We now consider the effect of a finite motion artefact, which manifests itself according to the assumptions developed in this section. We can then say that the *measured* signal will be given by equation [2.39], and that the ratio of ratios computed in the standard manner using a pair of artefact corrupted signals will be given by

$$\frac{v_j(t,\lambda_1)_{AC}}{v_k(t,\lambda_2)_{AC}} = \frac{\ln(1+m(t)) + \dfrac{\gamma_j(\lambda_1)p_j(t)}{\beta_j(\lambda_1)}}{\ln(1+m(t)) + \dfrac{\gamma_k(\lambda_2)p_k(t)}{\beta_k(\lambda_2)}} = R\,. \qquad [2.43]$$

Equation [2.43] demonstrates that the presence of motion artefact can dominate the calculation, with the result that the ratio of ratios will tend to unity. This means that a calibrated pulse oximeter output would tend to a calibration dependent constant under conditions of motion artefact, regardless of the true oxygen saturation. Because the degree of output corruption will therefore depend upon how different the true output is from this constant, it is extremely difficult to quantify the expected error. This may in part explain why some authors report erroneously high readings during periods of motion artefact whilst others report readings that are too low[56].

2.5 SUMMARY

A heuristic model for photoplethysmography has been introduced, which separates the received intensity into components from different optical paths. This model enables complex behaviour to be expressed in a general manner, the physical interpretation of which can be made context dependent. The model has been interpreted in terms of the Beer-Lambert law and an example situation modelled by diffusion theory, attributing both the received pulsations and the static parameters of the heuristic model with physical significance. It has also been demonstrated that this type of interpretation method can be used to aid physiological characterisation. Whilst these interpretations have placed the heuristic model in the context of two common physical models, it has been observed that it would be possible to attribute distinct physical models to the different optical paths. This would enable an increased level of flexibility and generality over homogenous modelling. It has been shown how a logarithmic transformation of the received intensity can re-normalise the observed pulsations with optical information in a manner that aids interpretation. This re-normalisation also results in signal dynamics that are independent of the source intensity or receiver characteristics, simplifying any subsequent analysis. It is

suggested that interpretation of these re-normalised PPG signals often may not need to consider the static signal component at all.

It has been highlighted that a major cause of motion artefact is dynamic variation in the optical coupling between the probe opto-electronics and the patient, an effect we have termed *probe-coupling* artefact (although it is noted that the effects of dynamic venous blood volume may be interpreted as artefact in the context of technologies such as pulse oximetry). The distinctive nature of this approach is that the effect of probe-coupling motion artefact has been modelled as a linear factor modulating the coupling coefficients of both static and pulsatile tissue. This factor is assumed to scale with the unmodulated coupling coefficients, but its functionality is unchanged. It is acknowledged that there may be other classes of motion artefact for which this assumption may be inaccurate, especially those that produce differential path length changes between source and detector by skewing of the geometry, or where a gross movement has shifted the entire field of view of the opto-electronics. Nevertheless the ensuant simplicity afforded by the non-linear measurement in removal of this probe-coupling artefact justifies an investigation of this approximation. It will be demonstrated later that a large proportion of artefacts can be observed to be predominantly probe-coupling in nature. It is also worth noting that this approximation is more general than the assumption of additive artefact, which has received serious attention[63-66,71].

The effect of probe-coupling artefact on pulse oximeter output has also been explored, demonstrating that under conditions of artefact the output will tend to a calibration dependent constant, making it extremely difficult to quantify the expected degree of error in pulse oximeter output.

3. ARTEFACT REDUCTION

Deterministic removal of the modelled form of artefact can be achieved by utilising an additional measurement channel with a source that exhibits contrasting anatomical absorption characteristics, with non-linear transduction simplifying the inversion. Subsequent analysis of the resultant, or equalised, signals demonstrates simplifications that can be used to aid interpretation. A new method for pulse oximetry having inherent insensitivity to artefact may be developed from these equalised signals, and it will be shown that this reformulation also possesses increased generality and enables an important engineering simplification in implementation. By addressing the theoretical and practical issues of implementation, an artefact-reducing photoplethysmographic system can then be developed to explore the performance and scope of this methodology.

3.1 MOTION ARTEFACT EQUALISATION

Motion artefact has been modelled as a linear factor modulating the coupling coefficients of static and pulsatile tissue, an effect we have termed probe-coupling motion artefact. It has been demonstrated that the use of a non-linear receiver characteristic leads to a separation of probe-coupling motion artefact into an additional term that is independent of the source intensity or receiver characteristics (equation [2.39]). Removal of probe-coupling artefact can therefore be achieved by generating two independent measures of the transformed intensity described by equation [2.39], with the use of two coincident light sources, and performing a subtraction of the two logarithmic intensity measures,

$$v_{jk}^{eq}(t,\lambda,b=0) = v_j - v_k = v_0 \ln\left(\frac{R_j I_j}{R_k I_k}\right) + v_0 \ln\left(\frac{\beta_j(\lambda)+\gamma_j(\lambda)p_j(t)}{\beta_k(\lambda)+\gamma_k(\lambda)p_k(t)}\right), \qquad \textbf{[3.1]}$$

where j and k label the two light sources and it is assumed for the sake of simplicity that there is no dynamic venous blood volume. The resulting signal, $v_{jk}^{eq}(t,\lambda)$, is now insensitive to the presence of a finite motion artefact, $m(t)$, although both static and dynamic signal components have been modified by the subtraction.

3.1.1 Non-linear equalisation

Equation [3.1] can be linearised to first order in the relatively small pulsations,

$$v_{jk}^{eq}(t,\lambda,b=0) \approx v_0 \ln\left(\frac{R_j I_j \beta_j(\lambda)}{R_k I_k \beta_k(\lambda)}\right) + v_0 \frac{\gamma_j(\lambda)p_j(t)}{\beta_j(\lambda)} - v_0 \frac{\gamma_k(\lambda)p_k(t)}{\beta_k(\lambda)}, \qquad \textbf{[3.2]}$$

and can further be simplified by assuming an underlying origin for the dynamic path length changes that does not vary with wavelength (such as that used in section 2.2.2),

$$v_{jk}^{eq}(t,\lambda,b=0) \approx v_0 \ln\left(\frac{R_j I_j \beta_j(\lambda)}{R_k I_k \beta_k(\lambda)}\right) + v_0\left[\frac{\gamma_j(\lambda)}{\beta_j(\lambda)} - \frac{\gamma_k(\lambda)}{\beta_k(\lambda)}\right] p(t)\,. \qquad \textbf{[3.3]}$$

Equation [3.3] indicates that an approach based upon the use of two light sources of contrasting tissue coupling characteristics will provide the most effective isolation of the pulsatile signal. Significant contrast, for example, is obtained in the relative absorption by blood and other tissue constituents between the commonly used red and infrared sources. Alternatively, the far-infrared region displays marked contrast with respect to either red or infrared regions. Logarithmically re-normalised pulsations from sources of suitable wavelengths will therefore exhibit an extreme size difference, enabling equation [3.2] to be approximated by

$$v_{jk}^{eq}(t,\lambda,b=0) \approx v_0 \ln\left(\frac{R_j I_j \beta_j(\lambda)}{R_k I_k \beta_k(\lambda)}\right) + v_0 \frac{\gamma_j(\lambda) p_j(t)}{\beta_j(\lambda)}\,, \qquad \textbf{[3.4]}$$

where j labels the *signal* (or measurement) source and k labels the *control* source. Equation [3.4] shows that, with proper choice of source wavelengths and to a first order approximation, the dynamics resulting from this subtraction are independent of the control source and have the same form as the transformed dynamics of equation [2.28]. We can therefore say that the measurement signal has been *equalised* with respect to the control source. The dynamics of equation [3.4] may then be related back to a physical model (such as equation [2.7] or [2.21]) without any requirement to consider the tissue characteristics at the control source wavelength.

Although this non-linear equalisation methodology results in signal dynamics that are linear with respect to the modelled origin of the pulsations (equation [3.4]), the static component of the original signal has been greatly modified by the equalisation process. However, because of the inherent re-normalisation of the dynamic signal components, it may often be unnecessary to consider the modified static component when analysing equalised signals. This has been highlighted previously by

observation that specific interpretations of the heuristic PPG model can be performed using knowledge of the normalised pulsations, $\gamma p(t)/\beta$ (see section 2.2).

3.1.2 Practical considerations

There are a number of practical considerations, other than the choice of control source wavelength already discussed, which will determine the accuracy of probe coupling motion artefact reduction in any implementation of this technique.

3.1.2a Geometric source positioning

Excluding correct choice of source wavelengths, the most important factor influencing the accuracy of artefact reduction will be the geometric positioning of the signal and control sources. For our heuristic model of motion artefact to hold, the optical paths travelled by light from both signal and control sources must be identical, implying that the two sources must be coincident. In practice, the best that can be achieved is placing the two sources as close together as possible, such as the LED die bonding technique used to place the sources in pulse oximetry probes. Because the LEDs themselves are physically small and the angular pattern of light emitted is relatively broad, such techniques can give a very good approximation to truly coincident sources.

3.1.2b Source intensities

Another practical consideration is the intensity of the sources, which would usually be dynamically controlled to achieve the best possible signal to noise ratio. The intensities must be chosen to ensure that the total intensity at the receiver from the signal source is less than the intensity due to the control source. This is because a logarithmic subtraction is equivalent to the logarithm of a ratio, the polarity of which depends upon which component is the larger. Static (DC) components of nearly equal magnitude will therefore cause the logarithmic subtraction to be unstable, with the choice that the control source should be the larger component having advantages for practical implementation. The exact intensities are not critical, so that the source

power levels should be adjusted infrequently to provide adequate signal-to-noise levels but to ensure that light from the control source always results in a higher received intensity than that from the signal source.

3.1.2c Direct coupling

In our original definition of a heuristic PPG model (section 2.1), we expressed the possibility of a direct coupling of light between source and detector as additional coupling-coefficients, α_j (see equation [2.1]). In the case of appropriate design of transmission mode PPG sensors, it is reasonable to assume that the contribution to the total intensity from direct-coupled light is negligible, since the human tissue physically separates the source(s) and receiver. For reflection mode PPG, however, the source(s) and receiver must be placed in relatively close proximity, with no tissue barrier between them. It is common therefore for reflection mode sensors to exhibit a significant direct coupling of light, the exact values of α being determined by the probe construction. If we perform artefact equalisation in the presence of small direct-coupling of comparable size to the arterial pulsations, then equation [3.3] can be re-written as

$$v_{jk}^{eq}(t,\lambda,b=0) \approx v_0 \ln\left(\frac{R_j I_j \beta_j(\lambda)}{R_k I_k \beta_k(\lambda)}\right) + v_0\left[\frac{\gamma_j(\lambda)}{\beta_j(\lambda)} - \frac{\gamma_k(\lambda)}{\beta_k(\lambda)}\right] p(t) + \frac{v_0}{1+m(t)}\left[\frac{\alpha_j}{\beta_j(\lambda)} - \frac{\alpha_k}{\beta_k(\lambda)}\right], \quad \textbf{[3.5]}$$

where we have linearised with respect to both arterial pulsations and the direct-coupled light. Equation [3.5] demonstrates that the result of a finite direct-coupling is a residual artefact term that has not been completely removed by the equalisation process, and whose magnitude depends upon the difference between direct coupling relative to static tissue coupling at the two wavelengths. Since it is not possible to predict the tissue-dependent values of β, the best method for minimising this residual artefact term is to ensure that any direct coupling is much smaller than the static coupling from non-pulsatile tissue, $\alpha \langle\langle \beta$. This implies that any PPG sensor

intended for use with non-linear equalisation techniques should use optical screening between the opto-electronic components, specialised geometric arrangement, or other methods of reducing the direct coupling of light between source(s) and receiver.

3.1.2d Spectroscopic measurement

The use of non-linear artefact reduction in conjunction with spectral techniques, such as pulse oximetry, would require one additional source to act as the control. It has already been shown, however, that the resulting dynamics of an equalised signal are largely independent of the control source and have been re-normalised with optical information in a manner that may often make it unnecessary to consider the static signal components in the analysis. This technique is therefore well suited to spectrophotometric methods, especially when they rely on observation of an undistorted arterial pulsation.

3.1.3 Consequences for residual ambient artefact

We now investigate the consequences of the probe-coupling motion artefact equalisation on any residual ambient artefact that has not been completely removed by the electronic multiplexing and subtraction scheme. We can reintroduce the residual ambient artefact as a small additive wavelength-independent term (since it is an additional intensity that theoretically affects all signals identically) in the received intensity model of equation 2.2. If a residual ambient artefact, modelled by intensity $a(t)$, which has comparable size to the pulsations, $\gamma_j(\lambda)p(t)$, is added to the received intensity, then the resulting non-linear signal in the case of negligible motion artefact will be

$$\begin{aligned} v(t,\lambda) \approx v_0 \ln\left[1 + ZR_j(\lambda)I_j\left(\alpha_j + \beta_j(\lambda)\right)\right] \\ + v_0\left(\frac{\gamma_j(\lambda)p_j(t)}{\alpha_j + \beta_j(\lambda)} + \frac{R_a a(t)}{I_j R_j(\lambda)\left(\alpha_j + \beta_j(\lambda)\right)}\right), \end{aligned} \quad \textbf{[3.6]}$$

where R_a is the responsivity of the detector to the ambient lighting and may include any additional attenuation due to filters or electronic processing. Note that although the residual ambient artefact has been re-scaled by a factor $I_j R_j(\lambda)\left(\alpha_j + \beta_j(\lambda)\right)$ with respect to a linear representation of the signal, the ratio of residual ambient artefact to the pulsatile component, $R_a a(t)/R_j(\lambda)I_j(\lambda)\gamma_j(\lambda)p(t)$, remains identical for the linear and logarithmic situations, as we would expect.

3.1.3a The effects of equalisation

If we now perform the equalisation process with $a(t)$ added to both received intensities, then equation [3.4] becomes

$$v_{jk}^{eq}(t,\lambda,b=0,m=0) \approx v_0 \ln\left(\frac{R_j I_j \beta_j(\lambda)}{R_k I_k \beta_k(\lambda)}\right) + v_0\left(\frac{\gamma_j(\lambda)p_j(t)}{\beta_j(\lambda)} + \frac{R_a a(t)\left(R_k I_k \beta_k(\lambda) - R_j I_j \beta_j(\lambda)\right)}{R_k I_k \beta_k(\lambda) R_j I_j \beta_j(\lambda)}\right) \quad [3.7]$$

in the case of negligible direct coupling and motion artefact. The presence of finite motion artefact would serve to further reduce the ambient artefact by a factor $\left(1+m(t)\right)$. Equation [3.7] shows that the ratio of residual ambient artefact to pulsatile component has been reduced by a factor τ by the equalisation process, where

$$\tau = \frac{R_k I_k \beta_k(\lambda)}{R_k I_k \beta_k(\lambda) - R_j I_j \beta_j(\lambda)}. \quad [3.8]$$

The size of any residual ambient artefact will therefore be reduced, relative to the size of the transformed arterial pulsations, by the motion artefact equalisation, provided that

$$R_j I_j \beta_j(\lambda) \leq 2 R_k I_k \beta_k(\lambda). \quad [3.9]$$

3.1.3b Conclusions

The constraint of equation [3.9] implies that a suitable choice of source intensities, such that the static magnitude of the control signal is greater than one half of the magnitude of the signal of interest, will lead to effective reduction of any residual ambient artefact. In practice it is necessary to ensure that the static control signal is greater than any signals of interest so that the logarithmic subtraction is stable, automatically satisfying this condition. It is also worth noting that the reduction is most effective when the static signal components (linear or logarithmic) are roughly equal, as we would expect.

It has therefore been demonstrated that the equalisation process, being based upon a subtraction technique, will have the effect of reducing any residual ambient artefact. The equalisation process will never increase the size of the ambient artefact under realistic operating conditions, but the degree of reduction will depend upon the relative received intensities of the control and signal sources. In practice it would only be necessary to adjust the source levels for optimal ambient artefact performance once as the probe is attached to the patient, since the technique is insensitive to small changes in the static parameters.

3.2 A NEW METHOD FOR PULSE OXIMETRY HAVING INSENSITIVITY TO ARTEFACT

The introduction of a non-linearity in the observation of incident light intensity has enabled us to develop a method for motion artefact equalisation based upon the use of a control light source. It has further been demonstrated that the non-linearity performs re-scaling of the observed arterial pulsations, and that the static PPG signal is transformed in a non-linear manner by the equalisation process. However, careful consideration of the relative sizes of terms has indicated that a suitable choice of source wavelengths will result in observed pulsations that are largely independent of the control source. It is therefore possible to use a pair of equalised PPG signals to determine the arterial oxygen saturation, in a similar manner to that of conventional

pulse oximetry. The use of equalised PPG signals will result in a pulse oximeter that has a reduced sensitivity to errors caused by patient movement and exposure to ambient light, alleviating the clinical problems caused by artefact induced errors (see section 1.4.1).

3.2.1 Methodology

A pair of signal sources, at wavelengths usually used for pulse oximetry, can be combined with a single control source that displays a high degree of contrast in relative absorption by blood and static tissue with respect to both measurement wavelengths. The intensity due to all three sources may then be transformed by the non-linear characteristic of the receiver, normalising the arterial pulsations with optical information. Equalised signals from the pair of measurement sources can then be obtained by subtraction of the transformed control signal, resulting in the removal of probe-coupling motion artefact and the reduction of any residual ambient artefact. The commonly used red and infrared sources may both be equalised with respect to another wavelength that displays markedly reduced absorption in blood relative to the static tissue. Alternatively, the recently suggested use of a pair of near-infrared sources on either side of the isosbestic point for pulse oximetry[26] implies that a common control source could utilise the red region (although this may complicate the calibration scheme).

The Beer-Lambert law is usually used as a theoretical basis for the understanding of pulse oximetry, and has been introduced previously (section 1.3.2). Although its physical applicability is questionable[33,34], the empirical calibration of contemporary pulse oximeters is conceptually based on a Beer-Lambert formulation[33]. Other practical limitations to the accuracy and general applicability of the calibration have also been discussed, resulting in the observation that a large number of effects, from unconsidered absorbers in the blood to physical or physiological heterogeneity, can seriously degrade the calibrated accuracy (see section 1.3.5). Others[13,14] have sought to overcome these theoretical and practical problems by reformulating the conventional derivation with diffusion theory, attempting to widen the theoretical

applicability and introducing modified calibration schemes. In our reformulation of pulse oximetry we must therefore attempt to maintain the maximum degree of generality (and therefore applicability) possible.

3.2.2 A new method for pulse oximetry

Generalisation of the oxygen saturation derivation can be achieved by utilising the inherent separation of optical paths defined by our heuristic model. Coupling coefficients define optical coupling efficiencies that vary between the optical paths, enabling the possible application of distinct physical models for light transmission through the different paths.

The use of a common physical model for all optical paths has already been demonstrated, with examples taken from the Beer-Lambert Law and diffusion theory (section 2.2). In both these cases the underlying physical model was used to formulate an expression for total light passing through the patient's anatomy. A relatively small change in optical path length, which was assumed to be the origin of the observed pulsations, was then introduced and the resulting dynamic intensity linearised with respect to these small changes. The relative size of the static and dynamic signal components was then used to assign physical meaning to the coupling coefficients.

We can reverse this process, by initially assuming a physical model for the source of observed pulsations; in both the examples presented we found that a dynamic change in physical path length through blood (Beer-Lambert law) or the entire homogeneous physiology (diffusion theory) was justified. Because of their small size relative to the static signal, the observed pulsations may be linearised with respect to the modelled path length change. The interpretation of the ratio of coupling coefficients then becomes all that separates the underlying physical transmission models. In the case of the Beer-Lambert law the ratio is equal to the effective absorbance of the blood, whilst in the diffusion theory case it is a function of the static path length and a wavelength-dependent characteristic of the tissue. Because the receiver non-linearity re-normalises the observed pulsations with respect to this ratio, we can say that the

observed pulsations have been scaled by a factor that can be interpreted by the specific physical transmission model employed.

3.2.2a Separation of optical paths

This methodology can be applied to pulse oximetry, by using distinct physical models to interpret received intensity from different optical paths identified by the heuristic model. Because the observed pulsations are so small, we are more justified in the application of the Beer-Lambert law to the dynamic transmission through pulsatile tissue than the case of total transmission through the entire anatomy. Using the same assumptions as the conventional derivation of oxygen saturation (see section 1.3.2), we may linearise the small change in transmitted light due to the change in optical path length (see section 2.2.2), enabling us to write

$$p(t) = -\Delta z(t)\mu_{blood}(\lambda). \qquad \textbf{[3.10]}$$

Equation [3.10], which is a linearisation of the Beer-Lambert law, can now be used as a physical model for the pulsations. Note the difference between this physical model and the model introduced by equation [2.9], where $p(t)$ was simply expressed as a dynamic path length change. The reason for expressing the pulsations in the manner of equation [3.10] is that the Beer-Lambert law physical model for the small changes in transmitted light is now incorporated into $p(t)$. This means that there is no restraint on the model for transmission through non-pulsatile tissue, which may take any form desired. The ratio of coupling coefficients, γ/β, now characterises the effect of different physical models for transmission through static and pulsatile tissue, with a ratio of unity indicating a uniform model for all optical paths. The dynamic (AC) component of an equalised signal (equation [3.2]) can now be expressed in terms of this model,

$$v_{AC}^{eq}(t,\lambda,b=0) \approx -v_0\Delta z(t)\left[\frac{\gamma_j(\lambda)}{\beta_j(\lambda)}\mu_{blood}(\lambda_j) - \frac{\gamma_k(\lambda)}{\beta_k(\lambda)}\mu_{blood}(\lambda_k)\right]. \qquad \textbf{[3.11]}$$

3.2.2b Ratio of ratios

Computation of the ratio of ratios, and therefore ultimately calibrated oxygen saturation, can be achieved by equalising a pair of measurement signals at wavelengths λ_1 and λ_2 with respect to a common control signal at wavelength λ_3. The ratio of ratios can then be determined from the two instances of equation [3.11] at the two signal wavelengths,

$$R = \frac{\mu_{blood}(\lambda_1)}{\mu_{blood}(\lambda_2)} = \mathrm{T}_{12}\left\{\frac{v_{AC}^{eq}(t,\lambda_1)}{v_{AC}^{eq}(t,\lambda_2)} + \mathrm{T}_{32}\,\frac{\mu_{blood}(\lambda_3)}{\mu_{blood}(\lambda_2)}\left(1 - \frac{v_{AC}^{eq}(t,\lambda_1)}{v_{AC}^{eq}(t,\lambda_2)}\right)\right\}, \qquad \textbf{[3.12]}$$

where

$$\mathrm{T}_{12} = \frac{\gamma(\lambda_1)\beta(\lambda_2)}{\gamma(\lambda_2)\beta(\lambda_1)}, \quad \mathrm{T}_{32} = \frac{\gamma(\lambda_3)\beta(\lambda_2)}{\gamma(\lambda_2)\beta(\lambda_3)}. \qquad \textbf{[3.13]}$$

The factors T_{12} and T_{32} are also ratios of ratios, and should therefore be substantially independent of any geometric coupling condition or individual tissue type. These factors could in principle be determined analytically from a static transmission model, but in practice are more likely to be estimated experimentally during calibration. Equation [3.12] indicates that numerical estimation of R from a pair of equalised signals also requires knowledge of the relative absorption by blood at the control and one of the signal wavelengths, $\mu_{blood}(\lambda_3)/\mu_{blood}(\lambda_2)$. The use of a control wavelength that displays substantially reduced absorption by blood with respect to the signal wavelengths would imply that this ratio would be small and largely independent of the blood oxygen saturation. The assumption

$$\frac{\mu_{blood}(\lambda_3)}{\mu_{blood}(\lambda_2)} \approx const \qquad \textbf{[3.14]}$$

can then be used to introduce another calibration parameter that could be estimated theoretically, but whose precise value might be determined by calibration. Having

estimated the ratio of ratios using a pair of equalised signals and our new calibration factors, we can then calculate the oxygen saturation in the standard manner (see equation [1.9]). The relation of equation [3.12] represents a generalisation of the conventional Beer-Lambert formulation, in that the underlying model for transmission through static anatomy is not predefined. The result is that three extra factors have been introduced, which serve to modify the calculation of the ratio of ratios, and represent additional parameters for the empirical calibration process.

3.2.2c Source dependent simplification

It is also worth noting the case when the control wavelength is chosen so that the relative absorption of light from the control source by blood is negligible with respect to the relative absorption of the measurement wavelengths,

$$\gamma(\lambda_3)\beta(\lambda_2)\mu_{blood}(\lambda_3) \langle\langle \gamma(\lambda_2)\beta(\lambda_3)\mu_{blood}(\lambda_2). \qquad \textbf{[3.15]}$$

In that case we can approximate equation [3.12] as

$$R \approx \frac{v_{AC}^{eq}(t,\lambda_1)}{v_{AC}^{eq}(t,\lambda_2)} \mathrm{T}_{12}, \qquad \textbf{[3.16]}$$

and we now calculate the ratio of ratios using only one additional calibration parameter, T_{12}.

3.2.2d Conclusions

This new methodology has several key advantages over the conventional formulation:

1. The method is insensitive to probe-coupling motion artefact, the presence of which would otherwise cause the calculated oxygen saturation to tend towards a calibration dependent value. In addition, the resulting PPG signals at the measurement wavelengths will also be equalised, enabling calculation of other

physiological parameters, such as heart rate, without interference from motion artefact.

2. Because the Beer-Lambert law has now only been used to model the relatively small changes in transmission through pulsatile tissue, this formulation results in a broader theoretical applicability.

3. The additional factors introduced provide extra degrees of freedom in empirical calibration schemes, conceptually enabling a wider calibration set. This may help to overcome some of the practical problems associated with calibration (see section 1.3.5).

4. A practical electronic implementation need only consider the AC (dynamic) signal components. This implies that the small AC signals may be electronically separated and gain may be applied to enhance the dynamic range of the measurement. This technique would improve the accuracy and resolution of the measurement over other methods, which either directly sample the total received signal or more commonly separate the signal components by subtracting an offset from the signal and applying a more limited gain[37] (see figure 3.1).

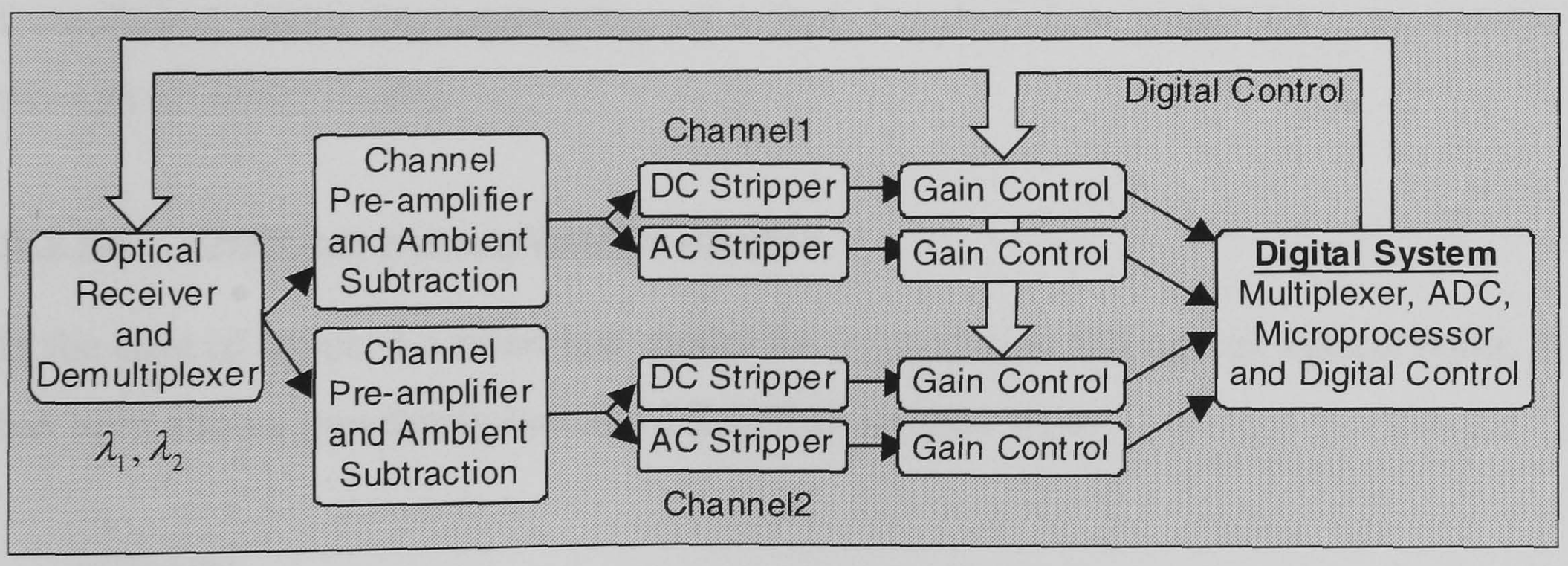


Figure 3.1 – Conventional pulse oximetry methodology

Because this technique is more general than the conventional formulation, to which it can be shown to reduce under conditions of a uniform Beer-Lambert transmission model, rigorous experimental verification is unnecessary to conceptually validate the

formulation. Figure 3.2 demonstrates the new methodology, which may be compared with the conventional methodology illustrated in figure 3.1.

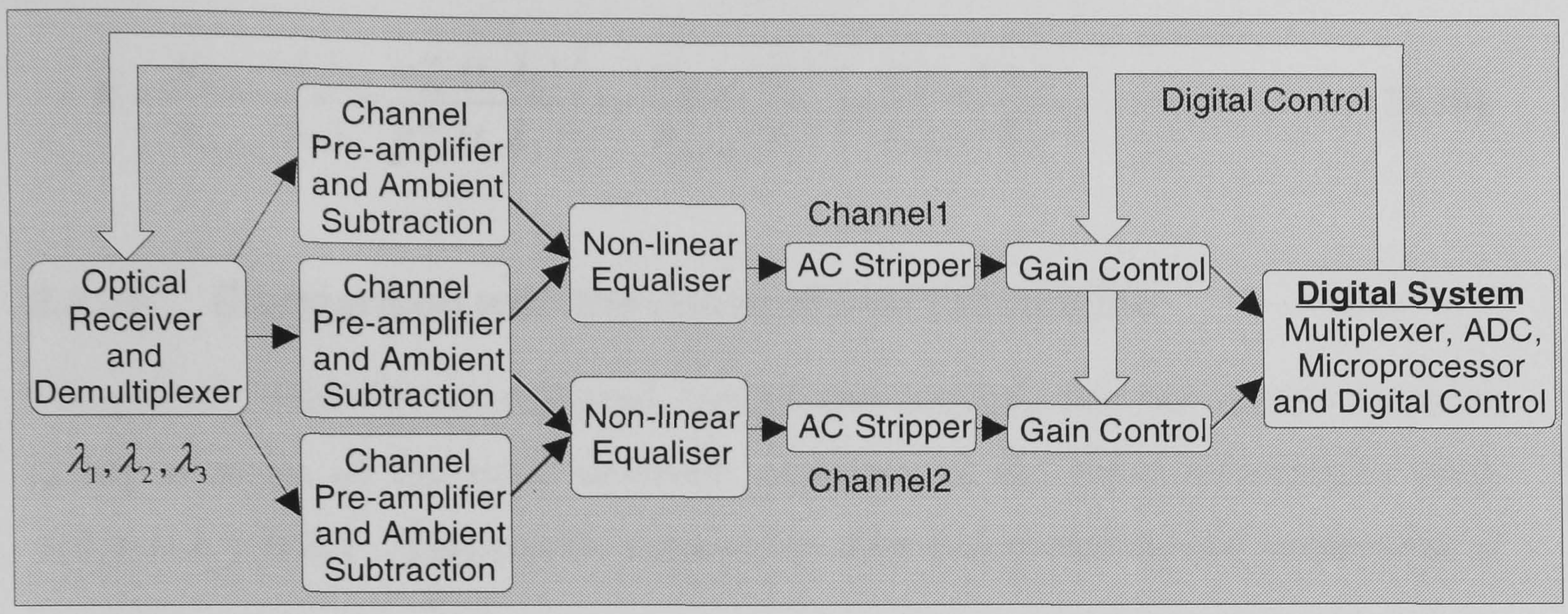


Figure 3.2 – Proposed new methodology for pulse oximetry

3.2.3 Relationship with the conventional formulation

It is worth investigating the relationship between equation [3.12] and the standard Beer-Lambert law derivation for calculating saturated oxygen. This exercise will demonstrate that the more general formulation will simplify to the conventional formulation, under the assumption of a Beer-Lambert law model for transmission through all optical paths.

3.2.3a Uniform transmission model

In the case of a Beer-Lambert law model for transmission through all optical paths, it has been shown previously (section 2.2.2) that we may write

$$\frac{\gamma(\lambda)p(t)}{\beta(\lambda)} = -\Delta z(t)\mu_{blood}(\lambda), \qquad [3.17]$$

and since the right hand side of this equation is just $p(t)$ in our general interpretation, we may say that the ratio of coupling coefficients must be unity and therefore that the factors T_{12} and T_{32} must also be unity. We can therefore simplify equation [3.12] to

express the ratio of ratios under the assumption of a uniform Beer-Lambert law transmission model,

$$R = \frac{\mu_{blood}(\lambda_1)}{\mu_{blood}(\lambda_2)} = \frac{v_{AC}^{eq}(t,\lambda_1)}{v_{AC}^{eq}(t,\lambda_2)}\left(1 - \frac{\mu_{blood}(\lambda_3)}{\mu_{blood}(\lambda_2)}\right) + \frac{\mu_{blood}(\lambda_3)}{\mu_{blood}(\lambda_2)}. \qquad [3.18]$$

3.2.3b Comparison with the conventional formulation

Comparison with the conventional formulation requires that we rewrite equation [3.18] in terms of the basic received intensities at the three wavelengths used, $I(\lambda_1), I(\lambda_2), I(\lambda_3)$. This can be achieved by first noting that the AC component of an equalised signal can be expressed as the time derivative of the subtraction of the transformed intensities at the measurement and control wavelengths,

$$v_{AC}^{eq}(t,\lambda) \equiv \frac{d \ln I(\lambda_j)}{dt} - \frac{d \ln I(\lambda_k)}{dt}. \qquad [3.19]$$

This enables us to write equation [3.18] in terms of the observed intensities and the relative absorbance of blood at the control and one measurement wavelength,

$$R = \frac{\dfrac{d \ln I(\lambda_1)}{dt} - \dfrac{d \ln I(\lambda_3)}{dt}}{\dfrac{d \ln I(\lambda_2)}{dt} - \dfrac{d \ln I(\lambda_3)}{dt}}\left(1 - \frac{\mu_{blood}(\lambda_3)}{\mu_{blood}(\lambda_2)}\right) + \frac{\mu_{blood}(\lambda_3)}{\mu_{blood}(\lambda_2)}. \qquad [3.20]$$

We can now invoke the conventional Beer-Lambert interpretation (see equation [1.8]) to express the ratio $\mu_{blood}(\lambda_3)/\mu_{blood}(\lambda_2)$ in terms of the received intensities, by defining another ratio of ratios,

$$R' = \frac{\dfrac{d \ln I(\lambda_3)}{dt}}{\dfrac{d \ln I(\lambda_2)}{dt}} = \frac{\mu_{blood}(\lambda_3)}{\mu_{blood}(\lambda_2)}, \qquad [3.21]$$

which allows us to simplify equation [3.20] to

$$R = \frac{\dfrac{d \ln I(\lambda_1)}{dt}}{\dfrac{d \ln I(\lambda_2)}{dt}}. \qquad [3.22]$$

Comparison of equation [3.22] and equation [1.8] indicates that the two methods are identical and that the general formulation does indeed simplify to the conventional algorithm under the assumption of a homogeneous Beer-Lambert transmission model. This is important in conceptually validating the new methodology, by demonstrating its equivalence to the conventional formulation for specific values of the additional calibration parameters.

3.3 AN ARTEFACT REDUCING PPG SYSTEM

This section outlines the design of an artefact reducing PPG system that has been produced to verify and explore the performance and applicability of the heuristic artefact model and subsequent artefact reduction. Electronic isolation of the arterial PPG signal is a delicate measurement to perform, and the overall effectiveness of any practical system must be a function of the specific signal processing utilised. It is for this reason that an objective test of the artefact reduction should provide a comparison between equalised and unequalised signals that have undergone identical signal processing. The design introduced here comprises a dual channel PPG system that employs both high speed multiplexing, high precision and bandwidth amplification and high resolution digital sampling techniques to obtain a high quality source of PPG signals. One channel of this system will act as the measurement source, which can then be equalised with the second, control source. An unbiased comparison of the measurement signal before and after equalisation can then be used to study the performance of the artefact reduction process. A detailed analysis of the electronic designs described in this section is presented in Appendix I.

3.3.1 Conventional PPG system

Before any artefact equalisation can be performed, a high quality source of two PPG signals is required. Since a reasonably high contrast exists in blood and tissue between wavelengths in the red and infrared, the artefact reduction process can be performed using a standard pulse oximetry probe, with the infrared source providing the measurement signal and the red source the control signal. This has several advantages over the use of a custom probe; firstly the two sources are a pair of LED die which have been bonded to a common substrate and encapsulated in a clear package, thus ensuring that the signal and control source are as close together as possible. Secondly, the use of a transmission mode finger probe allows straightforward reconstruction of motion artefacts commonly encountered in clinical practice. Finally, the proliferation of pulse oximetry probes gives easy access to a variety of physical constructions and source placements.

The design of a dual channel arterial PPG system for pulse oximetry probes obviously shares its fundamental concept with the front-end of a commercial pulse oximeter. The pulse oximeter, however, does not require the bandwidth or resolution of a research system and is flawed for such study by the inclusion of undisclosed signal processing stages. It is therefore necessary to re-evaluate the existing technology and to produce a high specification design that achieves the maximum possible signal quality.

3.3.1a System overview

The functional block diagram of the custom dual channel PPG system is shown in figure 3.3. Timing for the entire circuit is controlled by an 8-bit micro-controller, which gives a high degree of stability and flexibility. LED drivers alternately illuminate each LED when instructed by the micro-controller, with the electrical current delivered being controlled by an external digital interface (see appendix I.3 for details). Received intensity is converted into an electrical signal by the PiN diode and associated pre-amplifier circuitry. This signal is then de-multiplexed by sample-and-hold circuitry, which produces separated signals representing illumination due to

both sources and a measure of the ambient intensity. The ambient signal is then subtracted from either source signal to reduce the effects of ambient lighting, resulting in a pair of unprocessed PPG signals due to each illuminating wavelength.

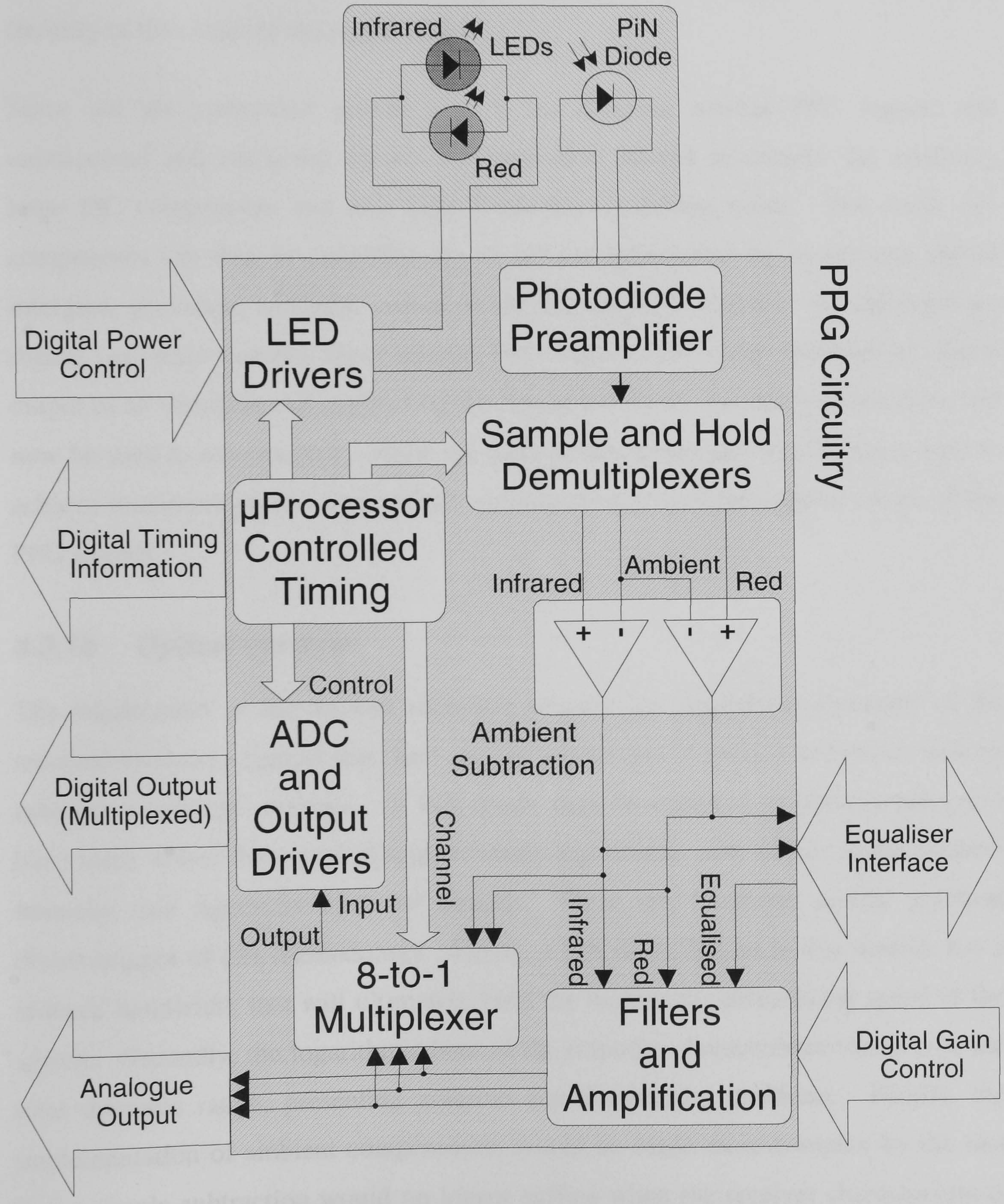


Figure 3.3 – System overview

It is at this stage that equalisation may be performed. In figure 3.3 the equalisation circuitry is shown located externally for the sake of clarity. The two unprocessed PPG signals act as inputs to the equalisation circuitry, which outputs an equalised version of the infrared signal. Inclusion of non-linear artefact equalisation into commercial PPG technologies could be achieved by the addition of equalisation circuitry at this stage of the signal chain.

Since we are concerned primarily with the dynamic arterial PPG signals, the unprocessed and equalised signals are band-pass filtered to remove the relatively large DC components and any high frequency switching noise. The small AC components can then be amplified by an amount determined by an external digital interface, providing analogue output of the arterial PPG signals. An analogue-to-digital converter samples these arterial PPG signals, providing multiplexed digital output of all channels and completing the digital interface. An external processor can now be used to automatically adjust the LED power levels and amplification level to achieve maximum performance, whilst simultaneously providing digital output of the PPG signals.

3.3.1b Optical receiver

The requirement of the artefact reduction process for logarithmic measures of the received intensity suggests that there may be advantages in using a transducer with an inherent non-linear response. A PiN diode may be operated in open-circuit (zero bias) mode where the output is approximately logarithmic with respect to the incident intensity (see Appendix I.1.1 for details). There are, however, several practical disadvantages of this methodology. Firstly, a PiN diode biased in this manner has a reduced bandwidth that will ultimately limit the maximum multiplexing speed of the system. Secondly, the logarithmic base of the response characteristic varies over the total dynamic range, presenting possible signal matching problems. Finally, the implementation of ambient compensation would be made more complex by the fact that a simple subtraction would no longer suffice when the receiver characteristic is non-linear. A PiN diode biased in the linear region is therefore the most suitable

choice, providing a high bandwidth and sensitivity, simplifying the ambient compensation and deferring the logarithmic transformation for a later signal processing stage. Maximal performance in linear-bias mode is achieved with a *transimpedance* circuit.

3.3.1c Sample-and-hold demultiplexers

Demultiplexing is usually performed with a simple analogue switch followed by a low-pass filter to remove the effects of switching noise from the signals. This has been replaced by sample-and-hold circuitry because such a system removes the need for filtering the signals at this early processing stage (see appendix I.1.2 for details). This is important for the equalisation process, which would be very sensitive to filter phase or gain errors that might otherwise be introduced. The sample-and-hold method of separating the signals does not require filters and therefore intrinsically results in better signal matching.

3.3.1d Timing considerations

So far we have assumed that the required signals should be generated by alternate source illumination and multiplexed measurement by a single receiver (see section 1.3.4). Although there are other ways to achieve simultaneous transillumination, such as wavelength division multiplexing using filtered receivers, the inherent signal matching obtained from a common multiplexed receiver offers many advantages. In addition, the use of a single transducer alleviates the problems of path length matching and simplifies the probe design, implying that a standard pulse oximetry probe may be used. Generation of the required timing signals is achieved with an 8-bit micro-controller, enabling generation of complex timing information whilst maintaining a high degree of flexibility (see Appendix I.1.2 for details).

This multiplexing technique has one major disadvantage, that the bandwidth of the signals is limited by the switching frequency. This is not a direct problem for the recovery of PPG signals themselves, whose bandwidth is significantly lower than practically achievable multiplexing speeds. The disadvantage is that the finite

bandwidth of signals may reduce the accuracy of artefact reduction, by effectively limiting the sensitivity of the system to a finite signal rate of change. The errors introduced into the ambient subtraction process by this limitation have been quantified previously (section 1.4.1b). The equalisation process, being conceptually based on a subtraction of signals, will be affected in a similar manner. Although the expected bandwidth of motion artefact is smaller than that of ambient artefact, the increased scale of motion artefact may result in a comparable signal rate of change. It is therefore necessary to maximise the switching frequency of the circuit, producing the maximum amount of information for both ambient subtraction and equalisation circuitry. The bandwidth of the PiN diode is the primary limiting factor to achievable multiplexing speed, with increased bandwidth competing with optical sensitivity. The practical implementation used a 5.05kHz switching frequency for each LED, representing a trade-off between adequate sensitivity and maximum speed.

3.3.1e Filtering and amplification

Linear-phase bandpass filters were used to limit the PPG dynamics to a spectral range extending from 0.75Hz to 5Hz (see appendix I.2). Because all ambient subtraction and motion artefact equalisation is performed before the filtering stage, there is no lack of accuracy associated with this bandwidth restriction. Instead, the filters merely serve to remove the DC component of the signals, enabling the application of enhanced gain, whilst reducing any high frequency switching noise, residual ambient artefact or electro-magnetic noise. The level of gain applied to these dynamics is determined by digital inputs, with a logarithmic relationship between the digital input word and the gain level (each bit is equivalent to 2dB), giving the circuit a wide dynamic range.

3.3.1f ADC and digital output

An 8-to-1 multiplexer provides input to the analogue-to-digital converter (ADC), allowing the high-speed digital output to sequentially step through several data channels (see appendix I.1.2 for details). In this way, a digital processor connected to the circuit may monitor the DC red and infrared signals as well as the filtered and

amplified dynamic signals. This is important because it is necessary to monitor the DC signal levels in order to correctly determine the necessary LED power and AC gain levels. Simultaneous monitoring of the equalised and unequalised infrared signal is also required to enable quantification of the level of artefact reduction achieved.

3.3.2 Artefact reduction methodology

Having obtained a fundamental source of PPG signals at a suitable pair of source wavelengths, it is necessary to logarithmically transform the signals and perform subsequent subtraction of the transformed control signal from the transformed measurement signal, therefore producing an equalised signal. Since we have utilised a linear receiver characteristic, the logarithmic transformation must be performed using a separate processing stage. It should be noted that transformation of the linear signals is performed as early as possible in the signal chain to avoid the problems caused by mismatched circuitry and amplifier phase errors. The equalised PPG signal is then filtered and amplified to isolate the arterial signal in an identical manner to the unequalised signal, enabling a fair comparison.

It is desirable in a system that aims to compare signal quality to minimise both the number and complexity of signal processing stages, therefore minimising the introduction of possible errors. This requirement is exacerbated in the current situation by the necessity of the equalisation process to compare signals, via the subtraction of transformed signals. Consider the effect of a slight mismatch in logarithmic transformation characteristics for the pair of signal sources, which may include digital sampling errors or variation in analogue characteristics. The error introduced may be insignificant under consideration of a single signal, but when a pair of transformed signals is subtracted the motion artefact will only be successfully removed if the transformation characteristics are identical. This matching requirement is made more stringent still by the extreme nature of the signal-to-noise ratio, with the arterial pulsations being much smaller than possible artefacts and expected static signal levels. It is therefore suggested that it is more critical to

accurately match the accuracy and characteristics of the logarithmic transformations prior to subtraction than it is to precisely define the transformation. This observation will allow us to identify a highly suitable method of performing the artefact equalisation.

3.3.3 Choice of technology

The most important decision to be made for successful implementation of an artefact reducing PPG system is the technology used to perform the non-linear transformations and subsequent subtraction. At this high level, the choice consists of either an all analogue electronic circuit designed specifically for the equalisation process or a custom or programmed digital circuit that samples the analogue signals and digitally computes the difference between their logarithms.

A digital system offers many practical advantages, including ease of design, flexibility and precise control of the non-linear characteristic. Unfortunately, any digital implementation is intrinsically restricted in performance by quantisation errors and the limited dynamic range of the analogue-to-digital converter.

3.3.3a Dynamic range

Any digital sampling system can only resolve the analogue input to within a specific tolerance defined by the circuitry resolution, thus dividing the input range of the sampler into a fixed number of equally spaced regions, with a higher resolution sampler having a larger number of regions. The most common samplers have resolutions of 8-bits (256 levels), 12-bits (4096 levels) or 16-bits (65536 levels), although higher resolutions are available. The practical resolution is very much less than this however, because we need to sample very small dynamic signals (the arterial PPG) superimposed on relatively large quasi-static DCs (the venous PPG). This dynamic range problem can be highlighted by consideration of a PPG signal whose arterial component is 1% of the total received signal, when sampled by a 12-bit converter. In the best case scenario, the sampler can then resolve the static portion of the signal to around 0.025% accuracy (1 part in 4096), whilst the small dynamic

signal can only be resolved to around 2.5% accuracy (1 part in 40). A realistic situation might place the total signal level at around 25% of the sampler input range, resulting in static and dynamic accuracy of 0.1% and 10% respectively. The need to monitor a large range of static signals together with the very small pulsations implies that any digital equalisation system should employ the highest resolution converter available.

3.3.3b Quantisation errors

The quantisation problem for a digital system is further compounded by the non-linear characteristic that must be produced. The effect of this non-linearity is that the output sensitivity to input quantisation errors is dependent on the absolute signal level, because the regular distribution of inputs regions are distorted at the output by the non-linearity (see figure 3.4). Consider a digital system that samples an analogue input signal, I_{analogue}, and from that sample digitally computes the logarithm of the signal, O_{digital}. We can write the digital input signal as

$$I_{\text{digital}} = I_{\text{analogue}} + \Delta I_{\text{analogue}}, \qquad \textbf{[3.23]}$$

where $\Delta I_{\text{analogue}}$ represents the quantisation error due to sampling. The digital input signal is then used to compute the digital output,

$$O_{\text{digital}} = \ln\left(I_{\text{digital}}\right) = \ln\left(I_{\text{analogue}}\left(1+\frac{\Delta I_{\text{analogue}}}{I_{\text{analogue}}}\right)\right), \qquad \textbf{[3.24]}$$

and noting that the error due to sampling is small with respect to the total signal, we can approximate this to first order as

$$O_{\text{digital}} = \ln\left(I_{\text{analogue}}\right)+\frac{\Delta I_{\text{analogue}}}{I_{\text{analogue}}} = O_{\text{analogue}} + \frac{\Delta I_{\text{analogue}}}{I_{\text{analogue}}}, \qquad \textbf{[3.25]}$$

where $O_{analogue}$ is the desired logarithm of the analogue input. This analysis demonstrates that the error in output of a digital logarithmic characteristic not only depends on the sampling resolution at the input, but also on the size of the static signals. Both equation [3.25] and figure 3.4 demonstrate that the output is more sensitive to input error when the size of the input signal is small. This is equivalent to a reduction in the input resolution that depends upon the signal size. This is a severe problem in the context of artefact reduction, since the static signal levels may be different and dynamic due to the effects of motion artefact, resulting in an overall signal mismatch that is also dynamic and therefore impossible to distinguish from the desired dynamics following the subtraction. The relatively large scale of any motion artefact with respect to the small arterial pulsations implies that a dynamic error of this sort could easily become significant, even with the use of high-resolution sampling.

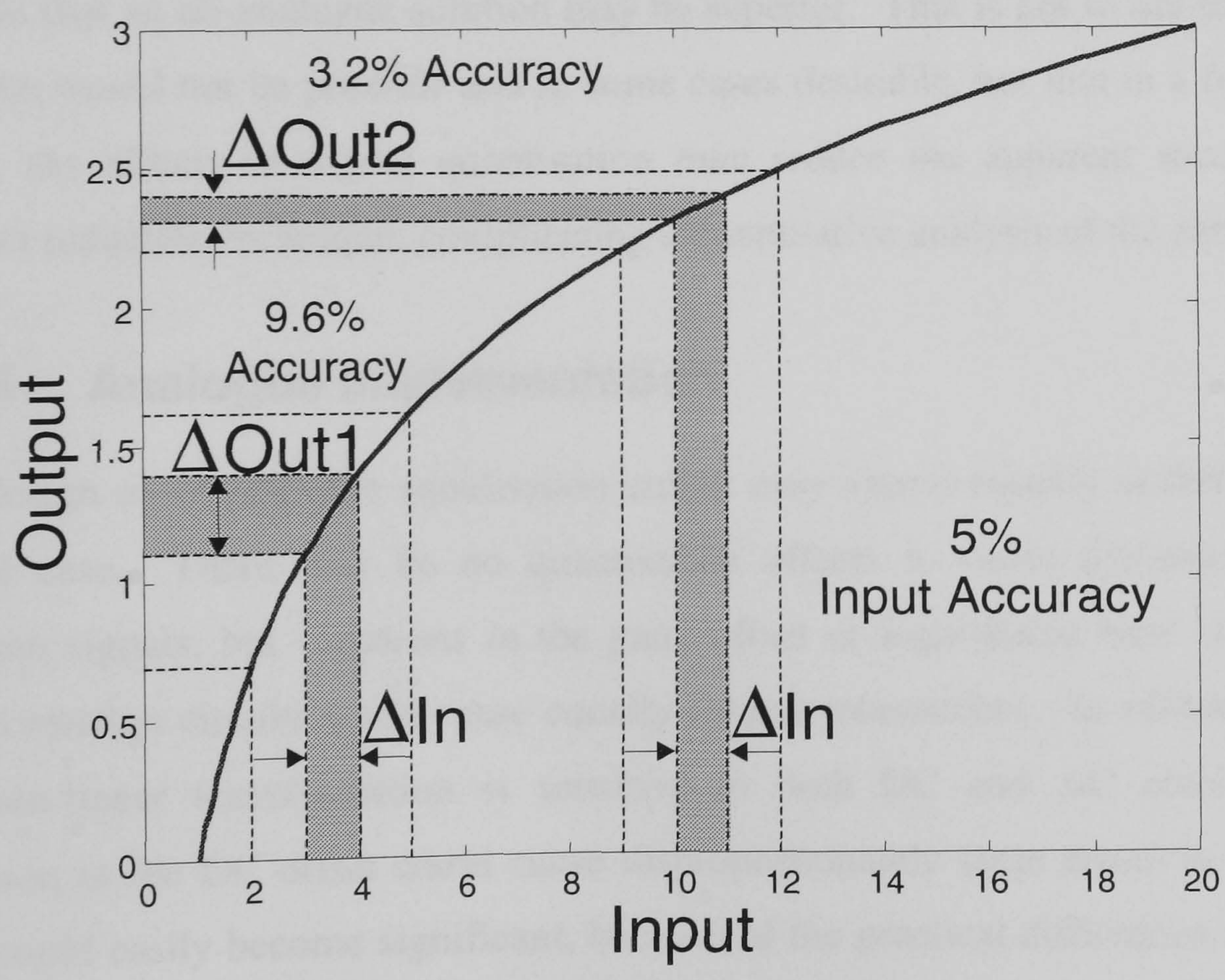


Figure 3.4 – Example of non-linear mapping of equal input regions and the effect on output accuracy

Whilst the problems of temporal quantisation usually associated with digitally sampled systems may also reduce the accuracy of equalisation, they do not dictate the choice of technology used to implement the equalisation process. This is because the source multiplexing that is used to achieve simultaneous illumination at the pair of source wavelengths is fundamentally slower than the relatively high speeds that can be obtained with contemporary analogue-to-digital converters. Therefore the information entering the equalisation circuit is already temporally quantised, and, providing that the digital sampler is synchronised with the source multiplexing, analogue-to-digital conversion will not exacerbate the problem.

3.3.3c Conclusions

The quantisation errors associated with a digital system, coupled with the wide dynamic range of the PPG signals and possible artefact, the non-linear nature of the necessary computation and the stringent requirements for signal matching together suggest that an all-analogue solution may be superior. That is not to say that a digital solution would not be possible and in some cases desirable, but that in a fundamental study, the effects of digital quantisation may reduce the apparent success of the artefact reduction technique, complicating a quantitative analysis of the performance.

3.3.4 Analogue implementation

The design of an analogue equalisation circuit may appear equally as difficult as the digital case. There may be no quantisation effects to cause dynamic mismatch between signals, but variations in the gain, offset or logarithmic base of individual transformation circuits could cause equally serious mismatches. In addition, because the non-linear transformation is sensitive to both DC and AC conditions, any variation in the DC offset could cause disproportionately large errors in the output. This could easily become significant, because of the practical difficulties involved in accurately maintaining the DC offset throughout several electronic processing stages. The solution to these practical problems is the use of a single circuit that performs the entire equalisation process in a single stage. This can be achieved by noting that

$$\ln(v_1)-\ln(v_2)=\ln\left(\frac{v_1}{v_2}\right), \qquad \textbf{[3.26]}$$

and constructing a single circuit to compute the logarithmic transformation of the ratio of the signals. Not only does this simplification lead to a convenient practical implementation, but also the use of a single circuit significantly reduces the possibility of matching errors. The practical offset, gain and logarithmic base of this circuit are no longer critical, provided that the characteristic is invariant across the operating range and the inputs to the circuit are matched. This method also maximises the practical dynamic range of the analogue circuit, which might otherwise be limited by impedance and signal-level matching between processing stages.

3.3.4a Non-linear circuitry

The classical method for producing a logarithmic amplifier is to utilise the non-linear relationship between the base-emitter voltage, V_{BE}, of a bipolar transistor and its collector current, I_C,

$$V_{BE} \approx \frac{KT}{q}\ln\left(\frac{I_C}{I_S}\right), \qquad \textbf{[3.27]}$$

where I_S is the reverse saturation current, K is Boltzman's constant (1.38×10^{-23}), T is the absolute temperature (K) and q is the charge of an electron (1.62×10^{-19} C). Converting an input voltage to a proportional current and passing it through the collector-emitter junction of a transistor results in a base-emitter voltage that is proportional to the logarithm of the input voltage. Figure 3.5 shows a simplified realisation based upon this concept.

A combination of a pair of such logarithmic converters may then be used to obtain the desired logarithmic ratio. The main disadvantage of this technique for application to the matched circuit required for artefact reduction is that computation of the logarithmic ratio still depends upon the use of two separate non-linear characteristics

(the transistors). For a matched logarithmic ratio, the two transistors must be both perfectly matched and isothermal. Practical logarithmic ratio amplifiers must therefore use identical transistors that are placed in thermal contact on a single silicon wafer and then laser trimmed for maximum accuracy. Even the matched transistors and expensive thin-film monolithic construction and ceramic encapsulation used in commercial integrated-circuit modules does not provide perfect matching between the two characteristics.

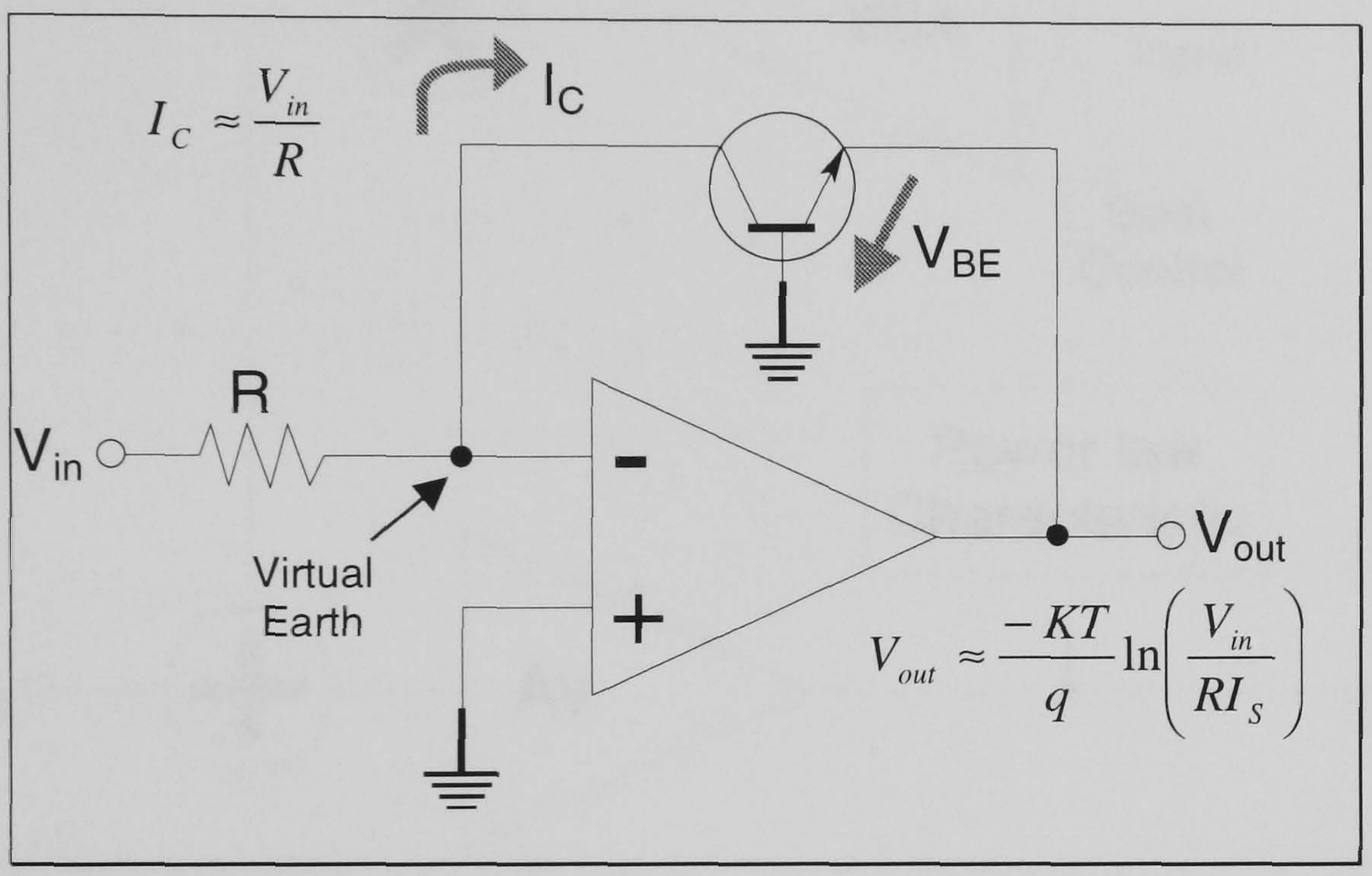


Figure 3.5 - Classical logarithmic amplifier

3.3.4b Feedback control

An alternate methodology for the production of a logarithmic ratio can be formulated by consideration of the inverse problem, namely the production of an exponential (or more generally a power law) characteristic. Inversion of equation [3.27] indicates that in the more usual common-emitter bipolar transistor biasing scheme, application of a base-emitter voltage will indeed result in an exponentially related collector current. It is only in the small signal domain that the common-emitter amplifier without emitter feedback may be considered to be linear; indeed it is usual in the construction of amplifiers to utilise feedback (emitter or otherwise) to linearise the amplifier output. Construction of a circuit with an exponential characteristic is therefore relatively straightforward to achieve. We can then use this exponential

characteristic to control the gain of a voltage-controlled amplifier (VCA) in the feedback loop of a high gain linear amplifier (see figure 3.6), resulting in a logarithmic output. Note that some common VCA integrated circuits already have an exponential characteristic (dB/V), eliminating the need to build a separate power law circuit.

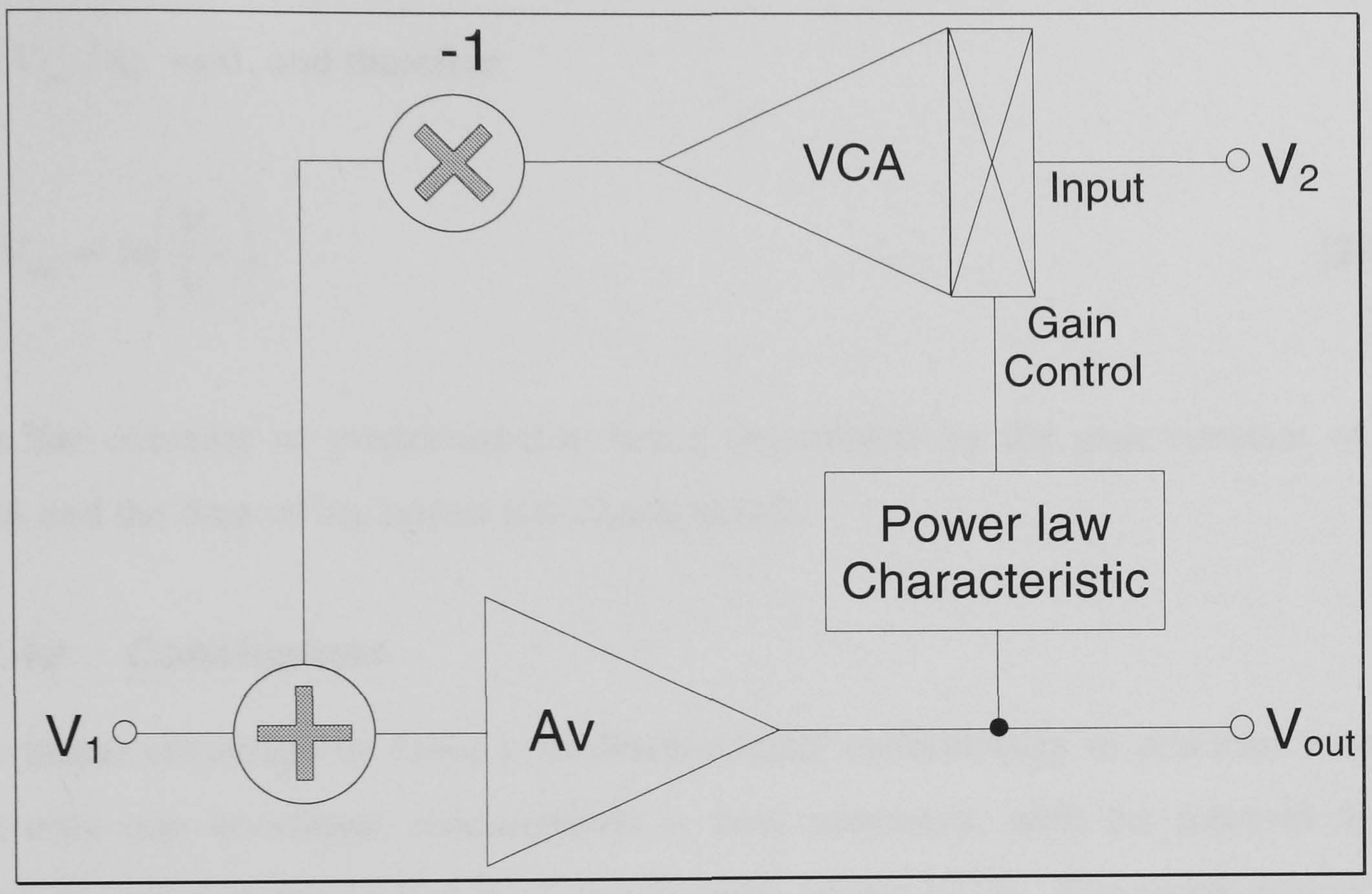


Figure 3.6 – Logarithmic ratio schematic

3.3.4c Feedback circuit analysis

Analysis of figure 3.6 demonstrates the principle of operation. The VCA multiplies its input, V_2, by a factor $ab^{V_{out}}$ where a is a gain constant of the VCA and b is the base of the power law characteristic. The output of the circuit can then be written in terms of the inputs,

$$V_{out} = A_V\left(V_1 - V_2 ab^{V_{out}}\right), \qquad \textbf{[3.28]}$$

where A_V is the large gain of the linear amplifier. Equation [3.28] can then be re-written as

$$V_{out} = \frac{1}{\ln(ab)} \ln\left(\frac{\frac{V_{out}}{A_V} - V_1}{-V_2} \right). \qquad [3.29]$$

If we now make the gain of the linear amplifier very large, $A_V \to \infty$, then we can say that $V_{out} / A_V \to 0$, and therefore

$$V_{out} \propto \ln\left(\frac{V_1}{V_2} \right), \qquad [3.30]$$

with the constant of proportionality being determined by the gain constant of the VCA and the base of the power law characteristic.

3.3.4d Conclusions

The major advantage of using a feedback-control methodology in practical terms is that only one non-linear characteristic is now necessary, with the inherent signal matching being dependent only on the linearity of the VCA. It is worth noting that this methodology would be unsuitable for a large proportion of existing applications of logarithmic ratio circuits, such as signal compression (or *companding*) and radio frequency modulation techniques, because of the bandwidth reduction that results from the use of a feedback technique. In the present application to artefact reduction, however, the expected bandwidth is comparatively low, whilst the extremely stringent requirement for matching the signal characteristics is adequately satisfied by using a single non-linearity. In addition, the range and availability of exponential VCA amplifiers and high gain operational amplifiers result in a relatively simple, compact and low-cost implementation, which is highly suitable for inclusion in the PPG signal chain. A detailed analysis of the actual equalisation circuitry used can be found in appendix I.4.

3.4 SUMMARY

It has been demonstrated that non-linear transformation of the received PPG signals results in the separation of probe-coupling motion artefact into a common additive term, the removal of which can be achieved by subtraction of a pair of transformed signals. It has further been shown that, with proper choice of source wavelengths, this difference results in a dynamic component that is independent of one wavelength and re-normalised with optical information. We have therefore proposed that a signal (or measurement) source may be equalised with respect to a control source of contrasting tissue coupling characteristics. The dynamic component of this equalised signal then contains all the necessary information for subsequent analysis and physical interpretation of the measurement PPG signal.

The practical considerations of implementing such an equalisation process have been highlighted, with emphasis placed upon the source positioning, wavelength selection and control of intensities. It has also been shown that the equalisation process will have the desirable effect of reducing any residual ambient artefact that has not been completely removed by the conventional technique. The degree of ambient reduction depends upon the static signal levels, suggesting further advantages to careful control of the source intensities. However, adjustment of the intensities need only be performed once after attaching the probe, since the method is insensitive to small changes in the static parameters.

A new method for pulse oximetry has also been introduced, which is conceptually based upon non-linear equalisation of a pair of measured PPG signals with respect to a third control signal. This re-formulation of pulse oximetry not only possesses inherent insensitivity to motion artefact, but also increased generality over the conventional formulation. This generality is achieved by restricting application of the Beer-Lambert law to the relatively small arterial pulsations, whilst expressing the possibility of a more general transmission model for static tissue as additional factors in the empirical calibration scheme. The validity of this method has been verified by

demonstrating that the calibration reduces to the conventional formulation under conditions of a uniform Beer-Lambert law transmission model for all optical paths.

The detailed issues of practical implementation of PPG artefact reduction have been addressed by the design of an electronic artefact reducing PPG system. This system encompasses all the usual functionality of a conventional PPG system, whilst employing a no-compromise methodology to reduce errors and improve system accuracy. In addition, the dual-wavelength system is capable of equalising one signal with respect to the other, enabling a fair comparison between a conventional signal and its equalised counterpart. A detailed analysis of the available technologies for implementation of the equalisation circuit has enabled us to identify a highly suitable methodology, which combines simplicity and straightforward construction with the greatest possible accuracy and signal matching. We are now in a position to examine the performance and scope of the equalisation methodology.

4. EXPERIMENTAL INVESTIGATION

Experiments with an artefact-reducing photoplethysmograph can be used to confirm several key assumptions and theoretical results, thereby verifying the validity of the non-linear approach. By definition and application of a quantitative artefact measure it is then possible to explore both the performance of artefact reduction and the scope of the underlying artefact model. In this way the performance and applicability of non-linear artefact reduction can be impartially assessed. It will be illustrated that a significant proportion of practical artefact conforms to this model, resulting in a high degree of artefact reduction in a wide range of clinically relevant situations.

4.1 CONFIRMATION OF THEORETICAL RESULTS

Experimental confirmation of theoretical results can be achieved with the practical artefact reducing PPG system developed in section 3.3. The use of a conventional transmission-mode pulse oximeter probe enables straightforward reconstruction of artefact producing effects encountered in a clinical environment, as well as providing suitable opto-electronics for the artefact reducing system. Direct comparison between equalised and unequalised signals can be used not only to investigate the performance of the system for artefact reduction, but also to verify several key theoretical assumptions and results. The results presented in this section are representative rather than exhaustive, their purpose being to demonstrate the important implications of observations made with the artefact reducing system. A more detailed and quantitative investigation of the performance of artefact reduction will be addressed in subsequent sections.

4.1.1 Self calibrating action

In section 2.3.2 it was observed that, because of the relative sizes of static and dynamic PPG signal components, the use of a non-linear receiver characteristic has the effect of re-normalising the observed dynamics with optical information in a manner that aids interpretation. One of the consequences of this observation is that, within the limits of practical implementation, the transformed dynamics are independent of the source intensity and transducer sensitivity. This hypothesis is straightforward to confirm experimentally, because the independence from source and transducer characteristics is maintained for an equalised signal (see section 3.1.1). This implies that changes in the source intensity should not affect the equalised signal, while the unequalised signal should scale proportionally.

Figure 4.1 demonstrates this self-calibrating action, by displaying the equalised and unequalised signals before, during and after a change in source intensity. The ten times reduction in infrared intensity was accomplished by digital control of the LED driver circuit, therefore ensuring constant physiological and optical coupling conditions throughout the test. Note that immediately following the change in

intensity both signals temporarily saturate the AC coupled amplifiers as the system adjusts to the new DC levels. The high degree of independence of the equalised signal from this intensity change not only verifies the self-calibration of these signals, but also highlights the reason that the proposed new method for pulse oximetry introduced in section 3.2 only needs to consider the dynamic signal components.

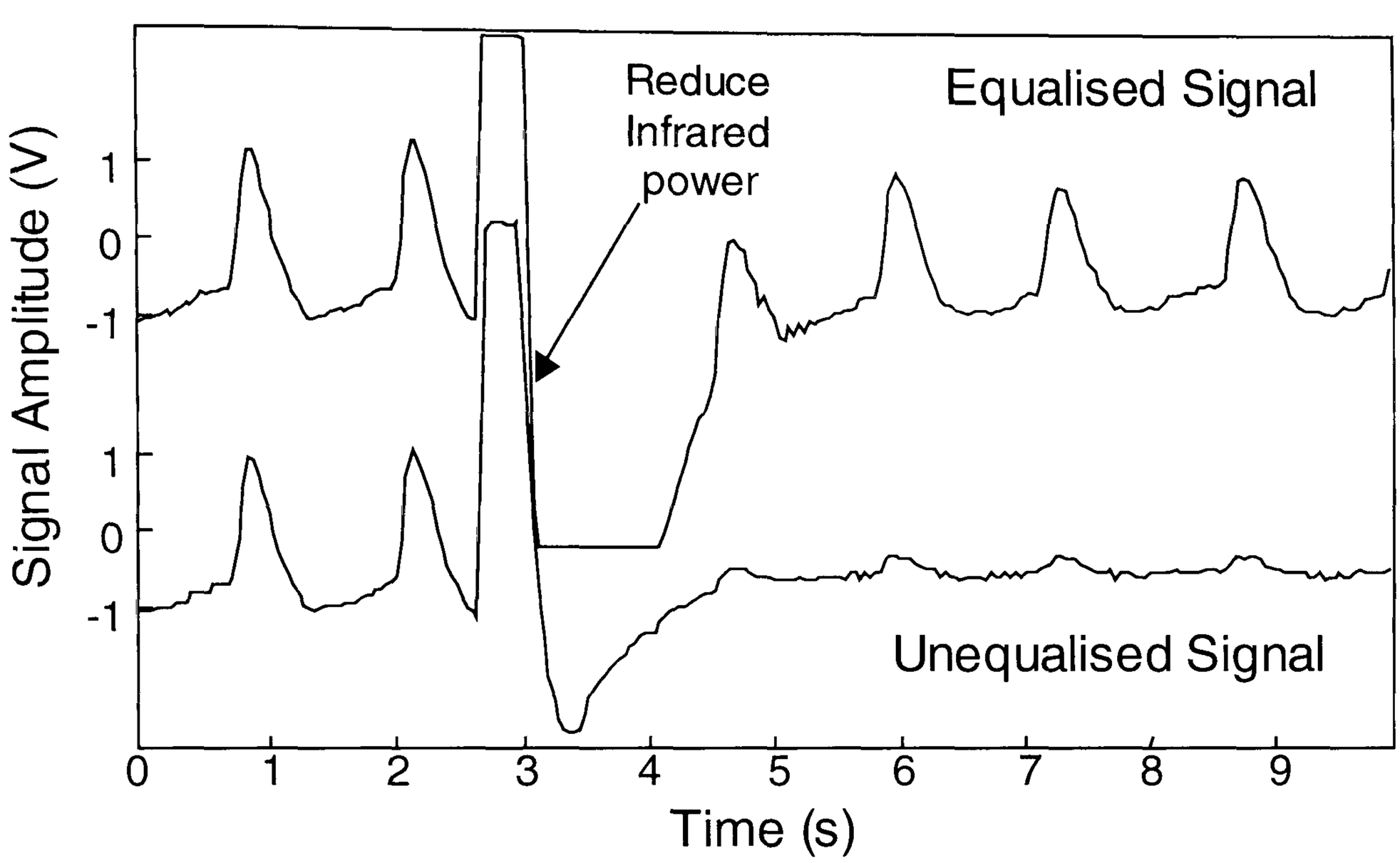


Figure 4.1 – Self calibrating action of equalised signal

4.1.2 Reduction of residual ambient artefact

The artefact reducing PPG system employs a much higher multiplexing speed than contemporary commercial PPG systems, which makes it inherently less susceptible to residual ambient artefact (see section 1.4.1b). In addition, the bandwidth limiting filters used to isolate the dynamic component serve to further reduce any high frequency ambient artefact from modulated light sources. It is still possible, however, to induce some residual ambient artefact by placing the finger probe in close proximity to an artificial light source. It was postulated in section 3.1.3 that the motion artefact reduction methodology should also reduce these ambient effects, depending on the relative intensity of the two sources. This can be verified by

comparison of the equalised and unequalised signals shown in figure 4.2, demonstrating the reduction of high frequency ambient artefact.

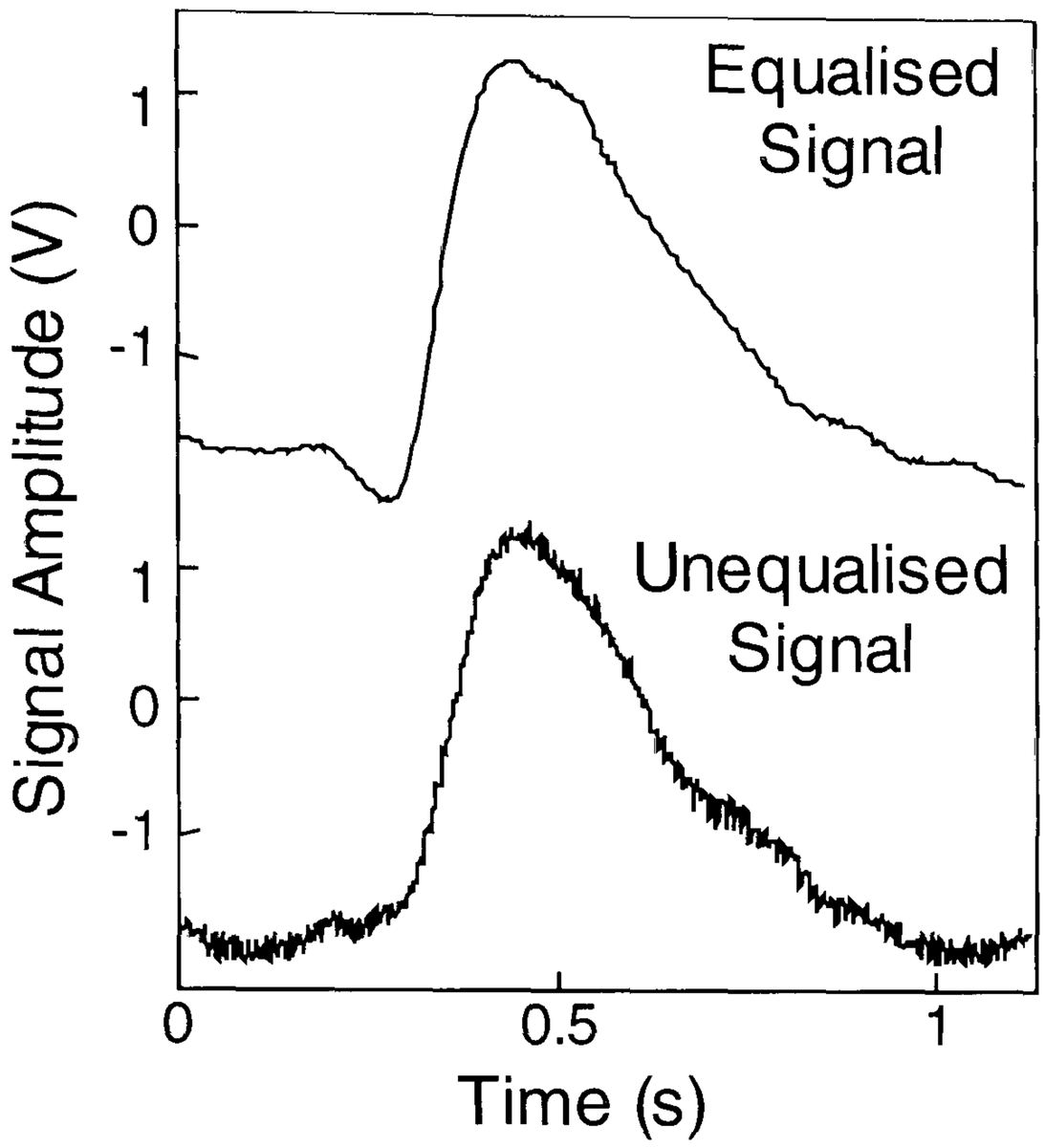


Figure 4.2 – Reduction of residual ambient artefact

The conditions necessary to achieve reduction of residual ambient artefact are that the static magnitude of the control signal must be greater than half the static magnitude of the signal of interest. Since the static control signal is necessarily greater than any signals of interest, this condition is automatically fulfilled and no further adjustment of the intensities was required to obtain figure 4.2.

4.1.3 Motion artefact equalisation

Whilst the quantitative performance of motion artefact reduction will be examined in detail later, it is worth highlighting several features of observed artefact reduction and their implications for confirmation of the theory. Figure 4.3 demonstrates the reduction of severe probe coupling artefact that has been intermittently induced by bending the finger periodically (at the distal-phalangic joint) within the probe. The success of artefact reduction is obvious, with the non-linear methodology recovering an uncorrupted PPG signal under conditions that have completely destroyed the unequalised signal. Visual identification of the artefact corruption is aided by the

relatively high frequency of the artefact relative to that of the uncorrupted PPG signal. The high degree of artefact reduction achieved serves to validate the basic assumptions about the expected nature of artefact corruption.

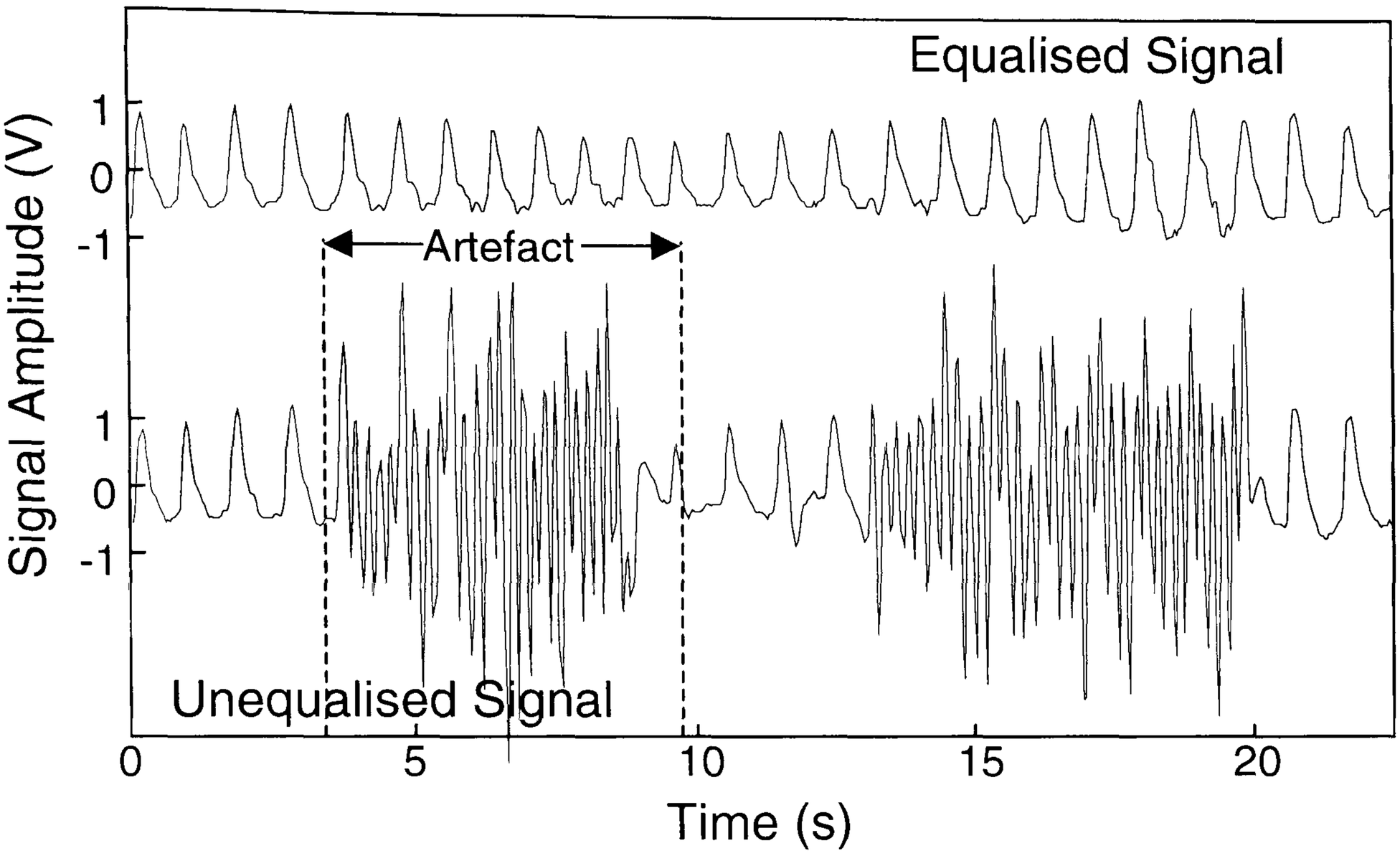


Figure 4.3 – Reduction of motion artefact

Another theoretical hypothesis is that motion artefact affects the PPG signals in a manner that does not depend upon the spectral characteristics of the artefact (see section 2.4.3). This is important because it enables us to control the frequency of induced artefact for the purpose of simplifying signal analysis, whilst maintaining generality. A broader spectrum of artefact than that shown in figure 4.3 was induced by pseudo-random waving movements of the hand, the results of which are shown in figure 4.4. The artefact induced in this way is not only of a lower frequency, but has a broader spectrum and size variability. It can be clearly seen from figure 4.4 that the artefact reduction methodology performs identically with broad-spectrum artefacts, confirming independence of the method from the frequency of artefact. It is worth noting that this type of broad-spectrum artefact displays significant spectral overlap with the uncorrupted PPG signal, implying that a spectrally based signal processing solution would be incapable of removing the artefact.

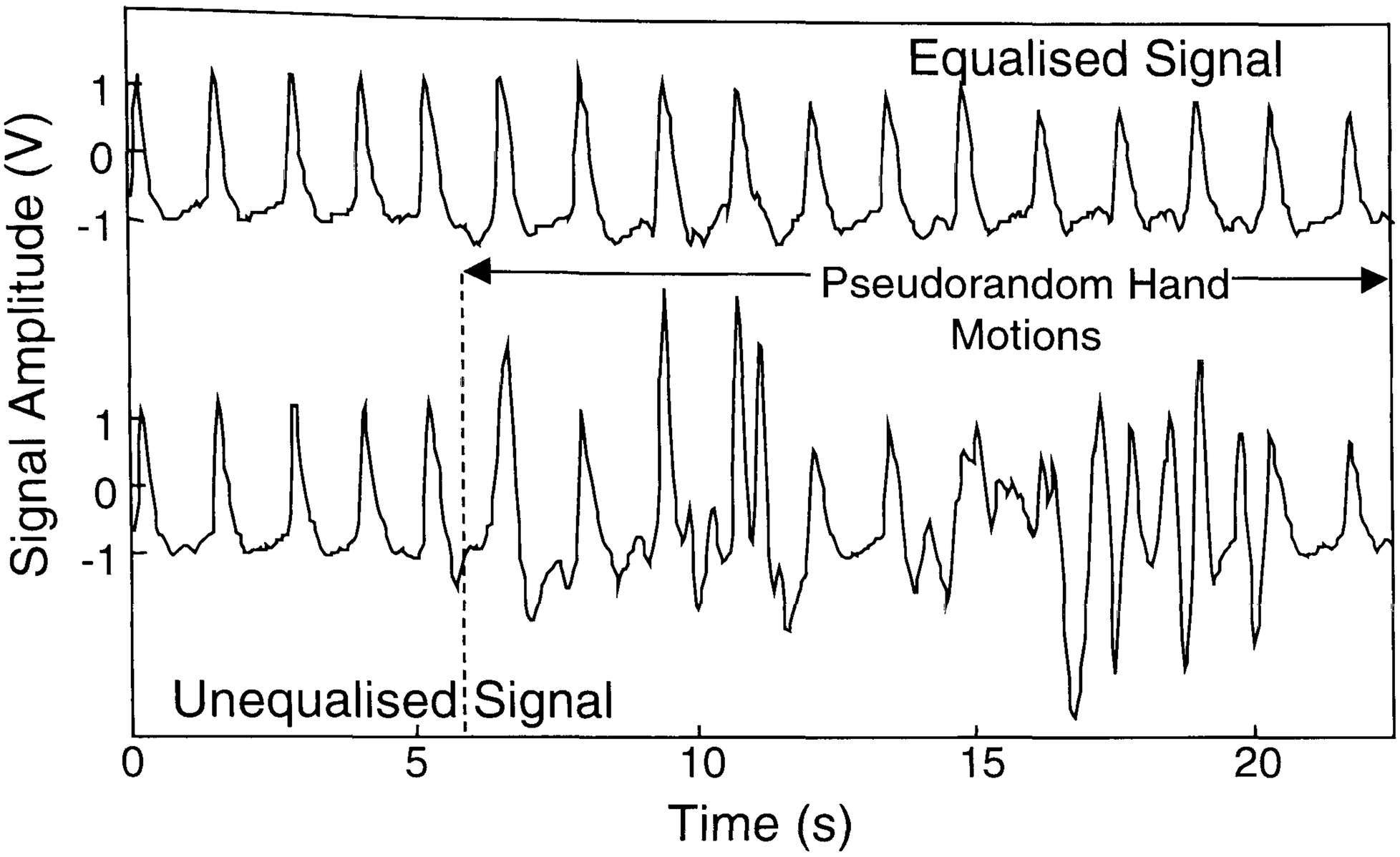


Figure 4.4 – Reduction of broadband motion artefact

4.2 QUANTITATIVE ARTEFACT MEASUREMENT

Precise and definitive classification of artefact and desired PPG signal components are an unachievable goal without additional information. Indeed, accurate identification of the artefact component of a PPG signal by *a-priori* knowledge would imply that a signal processing method could be used for artefact reduction. Artefact must therefore be strictly defined within the confines of the experimental methodology. This definition may employ comparison with an additional source of information or a more rigorous experimental protocol that intentionally induces artefact of known characteristics. Only in this way can we hope to quantify the severity of artefact on a given PPG signal, and therefore investigate the performance limits and applicability of the non-linear artefact reduction technique.

4.2.1 Artefact recognition techniques

Broadly speaking, there are two possibilities for the recognition of artefact corruption of a PPG signal. Some *a-priori* knowledge about the expected characteristics of the

clean PPG signal or the artefact, either implicit or artificially induced, can be used to separate the signal components. Alternatively, the expectation of a clean PPG signal can be verified against another independent transducer, either by direct comparison or as implied by interpretation of the signals. A quantitative comparison of artefact before and after the reduction process can then be used to investigate the performance of artefact reduction.

Because of the inherent commonality in the conceptual processes of artefact quantification and artefact reduction we can again examine extant solutions to artefact in the context of pulse oximetry (section 1.5), in an attempt to identify a suitable independent artefact recognition method. In fact, the commercial problems of artefact in pulse oximetry are more commonly tackled by the identification of artefact, and subsequent suppression of device output, than by its removal from the underlying PPG signals (see section 1.5.2). Careful control of experimental conditions, however, enables increased flexibility that is not afforded to the designers of commercial pulse oximetry equipment. Extension and modification of existing techniques is therefore possible, resulting in more sensitive and definitive recognition and quantification of artefact than would be otherwise possible.

4.2.1a Shape based recognition

The success of a shape based artefact recognition algorithm is entirely dependent on the accuracy and stability of the definition of expected shape. The pseudo-random nature of artefact makes it difficult to assume a-priori knowledge about its expected shape, precluding the direct identification or quantification of artefact in this manner. The more stable shape characteristics of a PPG signal do, however, provide some scope for identification of an arterial pulsation that has not been corrupted by artefact (see section 1.5.2). Identification can be based either on mathematical comparison of easily defined features[67], such as the gradient or zero crossings, or by quantifying the level to which a pulsation can be fitted to a mathematical model of an ideal pulsation[53]. Exclusion of uncorrupted pulsations from a given PPG signal leaves only pulsations that have been corrupted by artefact. The intrinsically digital

classification of pulsations can be made quantitative by defining artefact as the proportion of pulsations that have been corrupted within any given data set.

The use of artefact-recognition techniques dictate that the experimental conditions are controlled in such a way that artefacts having similar shapes to arterial PPG signals are not induced. Variation of the expected characteristics between individuals and experimental conditions must also be taken into account in order to obtain consistent results. Whilst these methods may work well during periods of extended and relatively severe artefacts, they are inherently limited in their sensitivity to both very severe and very mild artefacts. Severe artefacts cause all pulsations to register as corrupt, whilst mild artefact does not cause severe enough signal corruption for any artefact recognition to occur. Transient periods of artefact may also be misinterpreted by digital artefact recognition. In addition, digital classification of corrupt pulsations would result in a highly non-linear relationship between the artefact measure and the observed signal corruption and its effect on signal interpretation. Application of this method to artefact quantification for any given data set therefore requires that the conditions of both PPG signal and artefact are invariant with time, a situation that is difficult to achieve under practical physiological and experimental circumstances. Shape based techniques are therefore difficult to apply to the objective quantification of artefact reduction achieved by the non-linear methodology.

4.2.1b Verification by additional transducer

PPG signals may be verified by comparison to another independent transducer that does not suffer the effects of artefact (see section 1.5.2). Direct comparison between separate PPG sensors[68] conceptually provides a quantitative, linear and subject independent measure of the difference between the corrupted and uncorrupted signals, and therefore the artefact. Unfortunately, physiological heterogeneity between distinct locations implies that this technique would require very closely spaced sensors to be successful, making it very difficult to induce artefact in only one sensor. The problem would be further compounded by mismatches in the two sets of electronics required to drive the sensors, because a small amplifier phase mismatch

can result in signals that appear markedly different. A more practical technique is to validate the arterial pulsations by comparison of physiological parameters inferred by interpretation of the PPG signals and measured by independent means[56,69]. Secondary measurement of the pulse rate can be made by a variety of technologies, with the common ECG technique providing a stable beat-to-beat reading from a distinct location. Alternatively, groups of pulsations may be used to calculate other physiological parameters such as respiration rate or oxygen saturation, which can then be measured from alternate locations.

Consideration of the interpretation of PPG signals can therefore be used to classify arterial pulsations, or groups of pulsations, as corrupt or clean. An ensemble of pulsations may then be used to quantify the artefact in a data set, in the same way as the shape-based techniques. Again, it is this digital classification of pulsations that results in a non-linear relationship between this artefact quantification and the observed signal corruption, making it difficult to objectively judge the performance of artefact reduction.

4.2.1c Spectral techniques

The broad-spectrum nature of artefacts encountered in a practical environment mean that spectral filtering techniques will never be truly successful in either the removal or identification of realistic artefact. However, under experimental conditions it is a simple matter to ensure that artificially induced artefacts are spectrally separated from the predicted spectrum of arterial pulsations. Whilst a technique such as this would be entirely unsuitable for testing the performance of an artefact reduction technique based upon filtering or correlation cancellation, the non-linear artefact reduction methodology presented here is not dependent on the frequency of either the artefact or the arterial pulsations. In addition, the experimental artefact reducing system may be observed to perform identically regardless of the frequency of induced artefact, implying that restriction of the spectrum of induced artefact will continue to give objective results. It is therefore possible to utilise induced spectral separation of

signal and artefact to identify and quantify the artefact in a manner that would not be possible in other circumstances.

The use of spectral techniques offers many advantages for artefact quantification. Firstly, the analogue and continuous nature of such techniques can provide a quantitative artefact measure that is both linear with respect to the observed signal corruption and equally sensitive to mild and severe artefact. The quantification is itself continuous, enabling a true comparison of the magnitude of artefact between different sections of the same data set and removing the strict need for time invariant experimental conditions. Finally, the sensitivity of an artefact measure to signal variability between subjects is reduced in the frequency domain with respect to the time domain, implying that a spectral measurement will be more stable across a range of subjects and test conditions. Spectral techniques are therefore highly suitable for application to measuring the performance of the non-linear artefact reduction methodology.

4.2.2 Spectral artefact quantification

The simplest way to quantify PPG artefact spectrally would be to define a minimum frequency for the induced artefact, at a higher frequency than any expected signal component, which could then be separated by electronic filtering and measured by time integration. Unfortunately, this method would be over-simplistic in practice, because of a number of key observations:

1. The spectrum of the artefact extends down to the upper cut-off frequency for the arterial pulsations, even when the artefact fundamental frequency is much higher than the signal spectrum. This is because the multiplicative model of motion artefact (section 2.4) predicts mixing between the signal and artefact, resulting in broadband corruption.

2. The spectrum of the arterial signal will be dependent on the subjects pulse rate, and therefore the fundamental frequency of the pulsations. It is therefore

impossible to define an upper cut-off frequency for the expected signal without *a-priori* knowledge of the subject's heart rate.

3. The entire observable spectrum, for both signal and artefact, is necessarily narrow because of the filters used to isolate the arterial pulsations. Artefact of a relatively high frequency, apart from being difficult to generate, would be attenuated by the filters and degrade the accuracy of the quantification.

These restrictions imply that any filter that separates the artefact must have an adaptive cut-off frequency that depends on the subject's pulse rate and an extremely sharp transition from pass-band to stop-band. Adaptive filters of this kind may be realised using digital electronics. However, for the purposes of this study all analysis is performed on a computer, using previously stored data sets. This methodology removes the need to consider practical real-time implementation issues that could otherwise prejudice the results.

Development of a practical algorithm for the time dependant quantification of artefact is tackled in a number of steps. Firstly, uncorrupted PPG signals are obtained from a number of subjects under various testing conditions. The spectra of these signals is then used to estimate the expected bandwidth of a clean signal, relative to the fundamental frequency, or the pulse rate. Artefact may then be defined as the proportion of signal power that lies outside this expected bandwidth, discernible by Fourier transformation techniques. Data sets with increasingly severe artefact can then be analysed and the artefact quantification verified by correlation with the observed signal corruption. Spectral windowing techniques can then accurately separate the desired signal and artefact components, enabling time dependent quantification of the artefact. The separated components may also be used to demonstrate the ineffectiveness of pure signal processing techniques for artefact reduction, even under such carefully controlled experimental conditions. Finally, the average artefact measured in this manner may be compared between unequalised and equalised signals in order to judge the performance of the non-linear artefact reduction methodology.

4.2.2a PPG power spectra

Although any finite signal can be shown to have an infinite power spectrum, the use of data sets that are relatively long compared with the spectral information minimises any spectral distortion due to the finite experiment length. The detailed temporal shape of the arterial pulsations can also posses an infinite spectrum, although the contribution to the overall signal power drops dramatically after a few times the fundamental frequency. It is therefore not practical to analyse a data set and conclude an exact cut-off frequency for the PPG signal. More useful is a statistical measurement of the spectral power distribution, which can be used to define a practical cut-off frequency for the signal bandwidth, such that any information contained within higher-frequency components does not contribute significantly to the PPG signal. This can be achieved with the use of a Fourier transformation that is normalised with respect to the total spectral power. Integration of a region of this normalised spectrum therefore yields the proportion of signal power contained within that spectral range, enabling quantification of the contribution to the total signal.

PPG signals were recorded from a range of healthy adult volunteers, for a period of around 25 seconds per data set. Longer experiment times were not found to significantly alter the spectra, whilst increasing the possibility of physiological variability during the experiment. It was discovered that in most cases, 90 to 95% of the signal power was contained within a spectral region extending to three times the fundamental frequency. This leads to a convenient definition of artefact, as the proportion of signal power that has a frequency greater than three times the frequency of the signal fundamental, or subject's heart rate. The probability distribution of this artefact measurement when applied to the uncorrupted PPG data (figure 4.5) shows that, on average, artefact values of around 5 to 10% are to be expected for perfectly clean data, dependent on the characteristics of the specific signals considered.

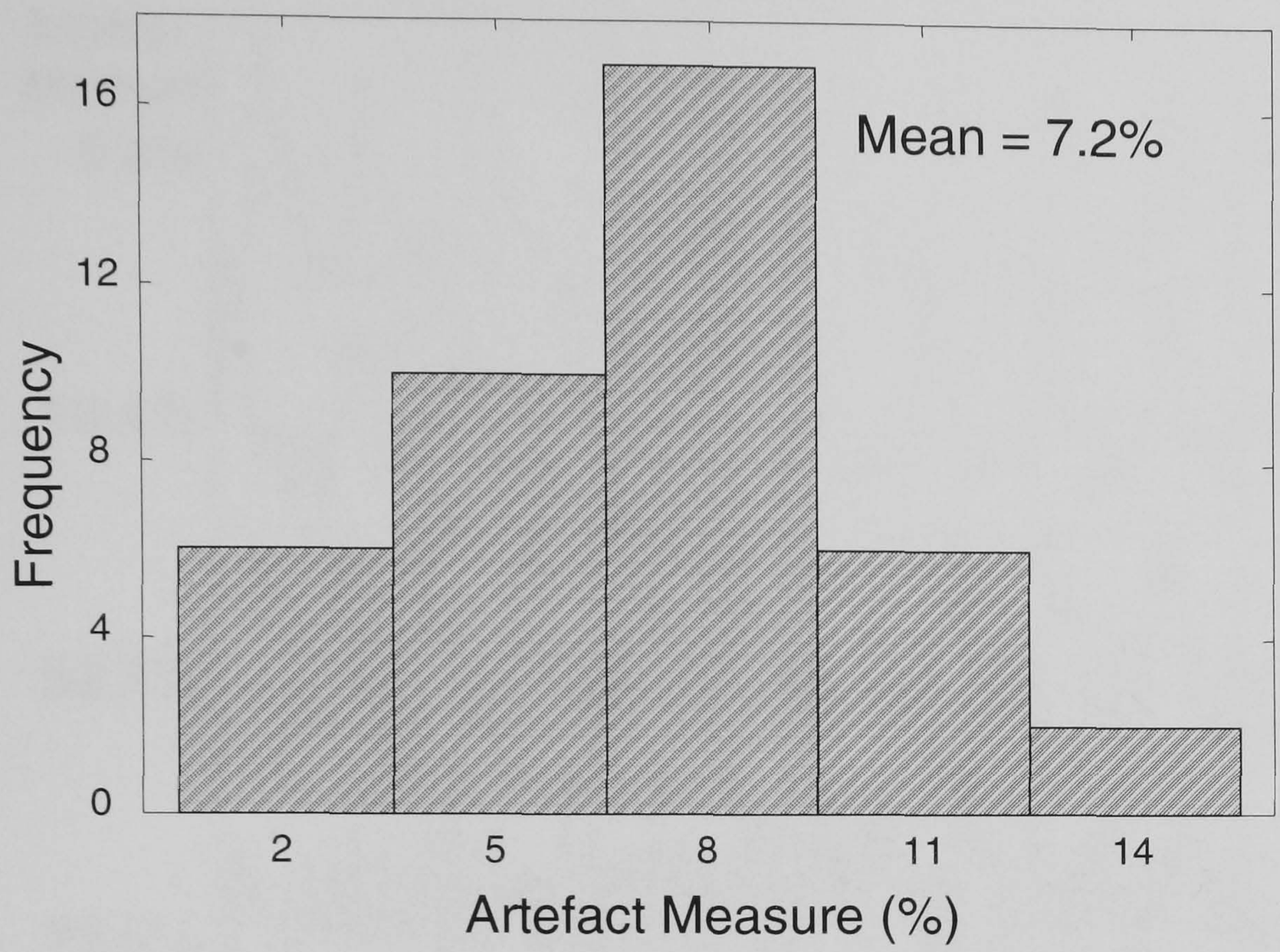


Figure 4.5 – Distribution of artefact measure of clean signals

This represents a minimum artefact magnitude, below which any high frequency artefacts cannot be distinguished from the intrinsic PPG signal components at those frequencies. If we induce artefacts of a frequency higher than three times the fundamental, then we expect the proportion of signal power in this spectral region to increase, thus increasing our measure of artefact.

4.2.2b Artefact quantification

A practical spectral artefact quantification has been defined as the proportion of total signal power that has a frequency higher than three times the frequency of the PPG signal fundamental. It has further been demonstrated that a lower limit on sensitivity exists, which is imposed by the infinite spectral bandwidth of any given artefact-free signal. It is now possible to analyse PPG signals from a single subject that have been corrupted to varying degrees by artificially induced artefacts of suitable frequency. In this way observation of the severity of signal corruption can be used to verify correlation between actual artefact and the spectral measurement (see figure 4.6).

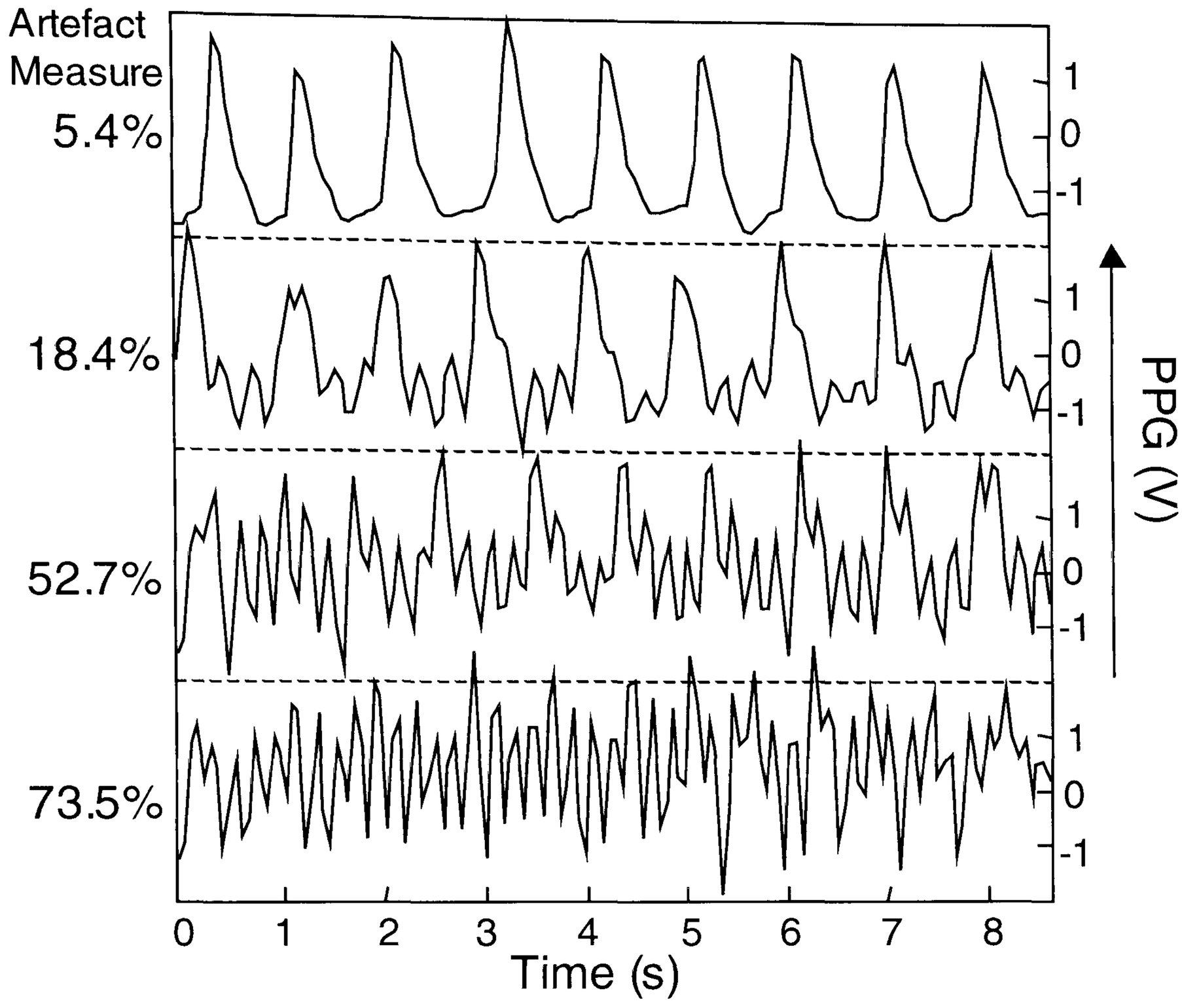


Figure 4.6 – Artefact quantification

It can be seen that this spectral measurement is both sensitive and increases monotonically with the observed signal corruption. A PPG signal with artefact measured to be 10% or under is considered to be clean and uncorrupted, whilst a signal with artefact measured around 20% is showing visible signs of artefact corruption. With artefact measured at 50% or more, the severity of corruption may make it impossible to correctly interpret the signals, and this level roughly represents the digital threshold for classification of individual pulsations. When the measured artefact is 70% or greater, the PPG signal has been completely destroyed and interpretation is now impossible.

4.2.3 Spectral windows

Having identified spectral regions that have been labelled as *signal* and *artefact*, it is now possible to spectrally separate these regions using spectral window techniques. This not only enables us to remove the region labelled artefact from the total signal,

therefore illustrating the scope of this definition, but also to isolate the region labelled artefact and perform a time dependent quantification. Figure 4.7 shows a 25-second sample from the artefact reducing PPG system. There are two distinct time periods during which severe artefacts have been artificially induced by periodic application of external pressure to the finger probe clip. The equalised version of the signal is also shown, demonstrating successful removal of the artefact. The unequalised signal has an artefact measure of 74.1%, whilst the equalised signal has a measure of 3.7%.

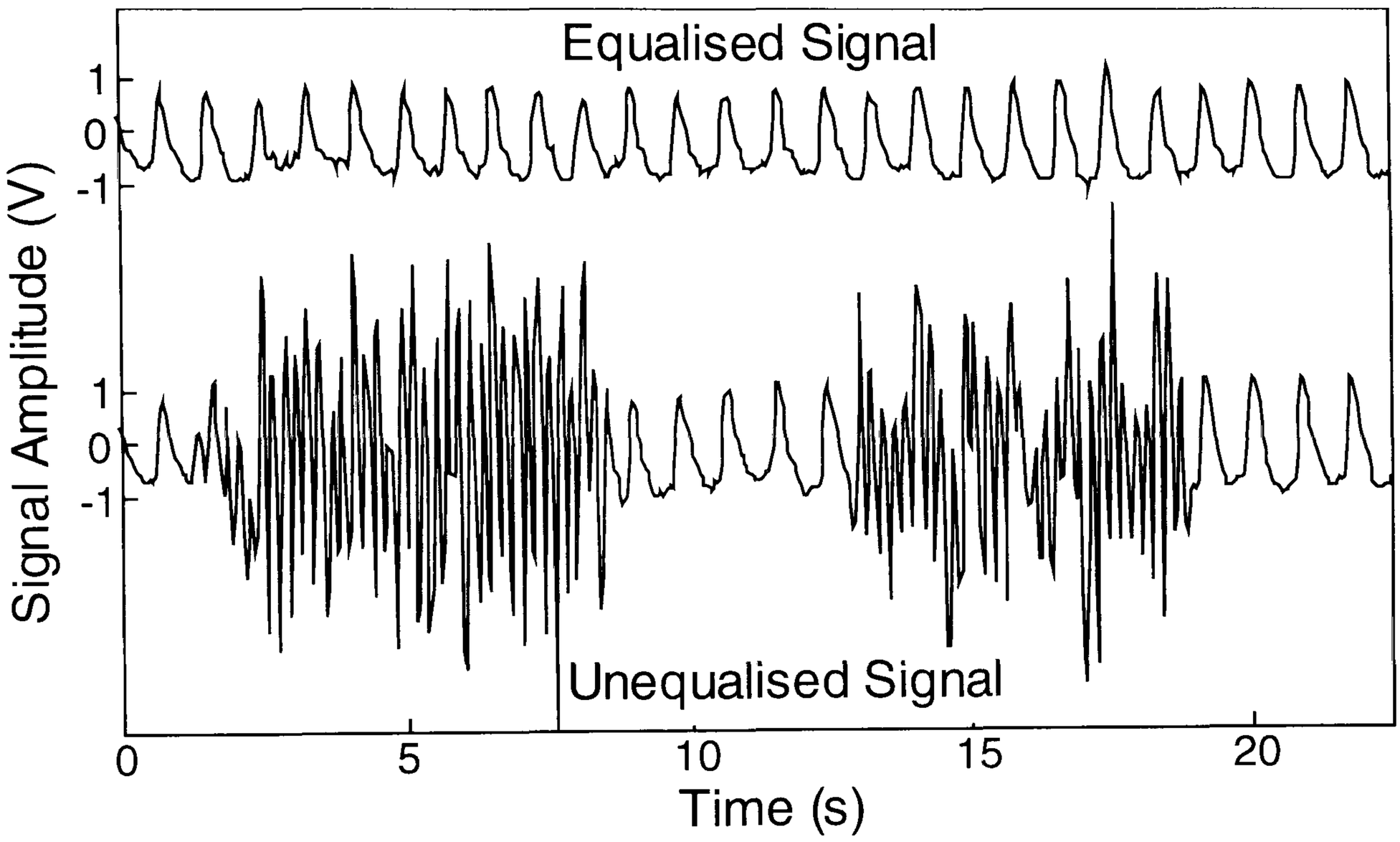


Figure 4.7 – Output of artefact reducing PPG system

Examination of the normalised power spectrum of the unequalised signal (figure 4.8) highlights the spectral separation between the signal fundamental and the artefact. The fundamental can either be identified as the lowest frequency peak in the spectrum, or by independent measurement of the subject's heart rate. The spectrum is divided into a signal and an artefact portion at a frequency boundary of three times this fundamental.

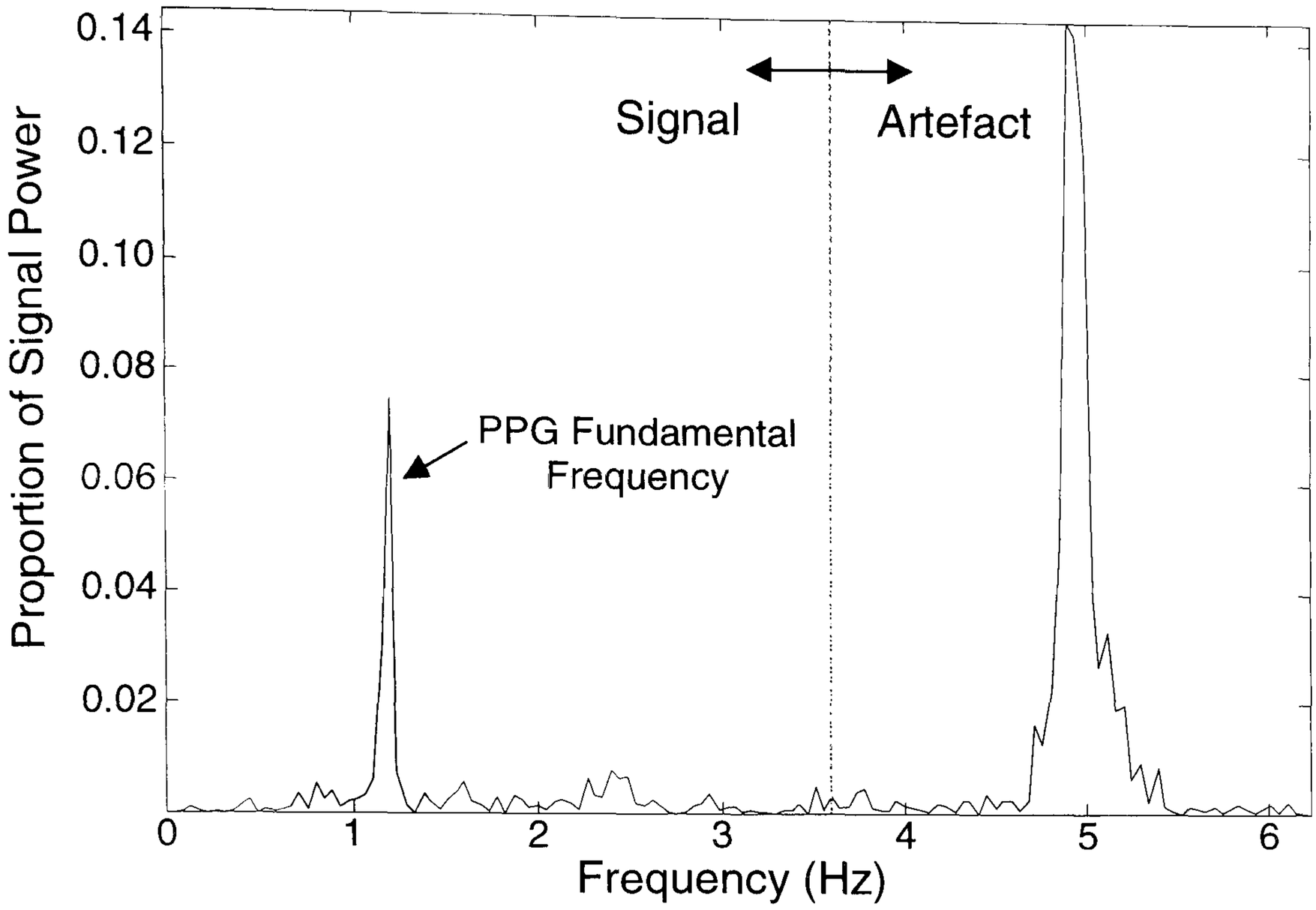


Figure 4.8 – Normalised power spectrum

4.2.3a Signal isolation

The portion of the spectrum labelled artefact can be removed from the signal with the use of an inverse Fourier transformation of only the desired signal portion. This technique can be conceptually compared to applying a perfect low pass filter to the received signal, with a *brick-wall* response and a cut-off frequency set to three times the subject's heart rate. Although such a perfect filter could not be realised in practice, this procedure serves to illustrate the theoretical performance limit of filtering techniques in general for removing artefact that is spectrally separated from the desired signals.

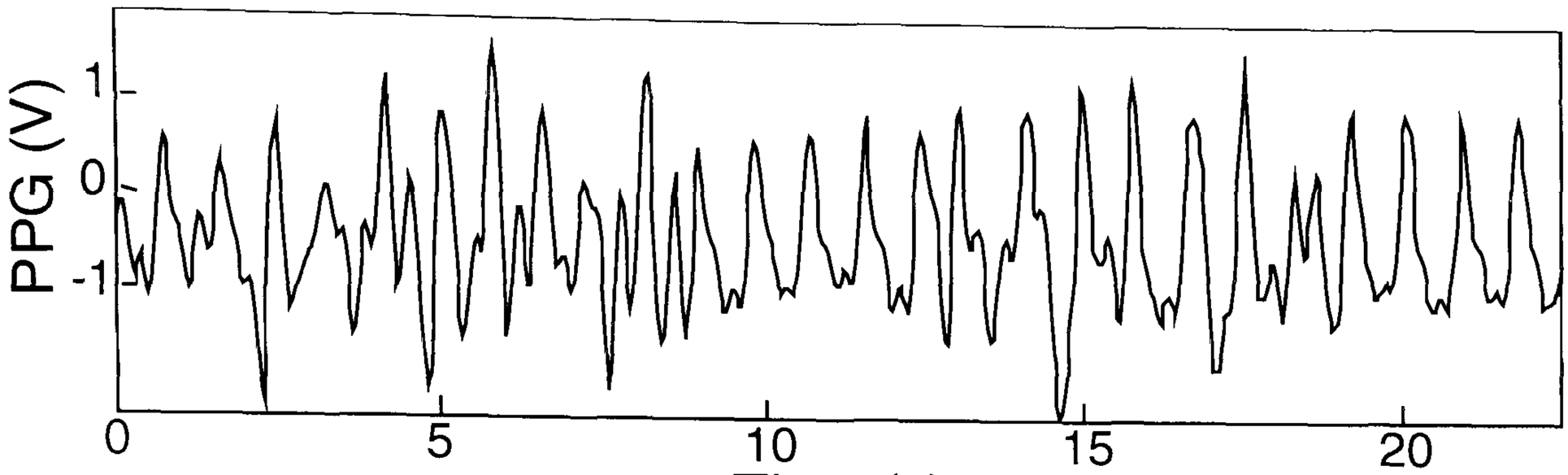


Figure 4.9 – Low pass filtered version of unequalised signal

The result of this inverse Fourier transformation can be seen in figure 4.9, which demonstrates that even a perfect filter could not completely remove the artefact. This is because of the multiplicative nature of motion artefact, which produces mixing between the PPG signal components and the artefact. Consider two discrete frequencies from separate PPG signal and artefact spectra, f_{PPG} and f_A. If the artefact is multiplicative, then the resulting signal due to these discrete frequencies will be given by

$$\cos(2\pi f_{PPG}t)\cos(2\pi f_A t)=\frac{1}{2}[\cos(2\pi\{f_A+f_{PPG}\}t)+\cos(2\pi\{f_A-f_{PPG}\}t)], \quad \textbf{[4.1]}$$

where the frequencies $(f_A + f_{PPG})$ and $(f_A - f_{PPG})$ will appear in the spectrum of an artefact-corrupted PPG signal. It is the presence of the difference frequencies, $(f_A - f_{PPG})$, which cause corruption of the continuous PPG spectrum to become broadband, even when the artefact is artificially induced to have a relatively high frequency. The inability of this technique to recover a clean PPG signal further verifies the assumption of multiplicative artefact (see section 2.4), since spectrally separated additive artefact would be completely removed by such a perfect filter. Artefact encountered in realistic situations will not have a pre-defined spectrum, further impeding the use of filters for its removal.

4.2.3b Artefact reconstruction

A similar technique to that used to remove the artefact portion of the signal can also be used to isolate it. Although examination of figure 4.9 indicates that the spectral separation between signal and artefact is not perfect, because of the multiplicative nature of motion artefact, the obvious success of the artefact quantification demonstrated by figure 4.6 suggests that artefact may still be measured in this way. The inverse Fourier transformation of the artefact portion of the spectrum of figure 4.8 can be seen in figure 4.10, which demonstrates how the magnitude of the high frequency components of the signal vary with time. This analysis is valuable in highlighting the origin of the lower limit of sensitivity on the artefact measure, by explicit demonstration of the higher frequency components of the uncorrupted portions of the signal. The two periods of severe artefact are clearly visible as a sharp increase in high frequency components of the signal.

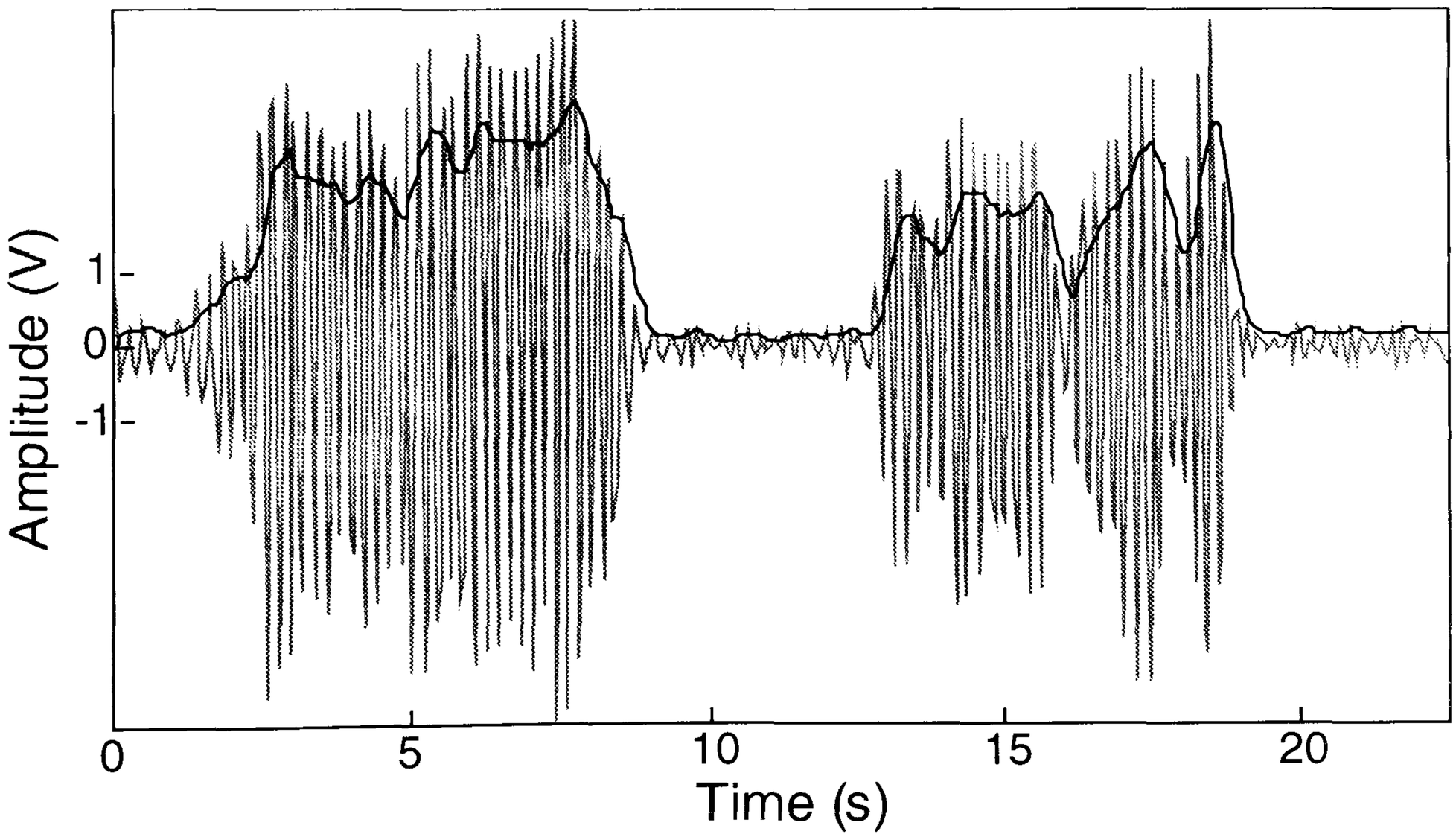


Figure 4.10 – Time dependent artefact measure

The envelope of the high frequency components, which is also shown in figure 4.10, can be used as an instantaneous measure of the high frequency magnitude, and therefore the magnitude of the artefact. In this way, it is possible to quantify the artefact in a time dependent manner. This can serve as a method for isolating periods of artefact from periods of uncorrupted signals. This technique will prove useful later

for ensuring that experimental conditions are approximately time invariant when investigating the scope and performance of the artefact reduction methodology.

4.3 QUANTITATIVE INVESTIGATION OF ARTEFACT REDUCTION

The fundamental lower limit on the sensitivity of artefact measurement precludes the absolute quantification of artefact reduction, unless both equalised and unequalised signals are severely corrupted. Nevertheless, quantification of artefact before and after the equalisation process can still be used as an indication of the effectiveness of the equalisation process under differing experimental conditions. In addition, consideration of the actual signal corruption produced by various bands of artefact measurement (figure 4.6) enables appropriate physical interpretation of numerical results. In this way, the scope of the non-linear artefact reduction methodology may be investigated experimentally.

4.3.1 Experimental methodology

In order to investigate the applicability of the artefact reduction technique, it is first necessary to define a series of practical conditions to be tested, which can be easily reconstructed experimentally. It is also necessary to perform a control experiment, whose results can be used as a basis for comparison with other, more complex situations. In this case, the most appropriate control experiment is the generation of pure probe-coupling motion artefact. The performance of artefact reduction achieved under conditions of pure probe-coupling artefact is a good indicator of the validity of models and assumptions used in obtaining an inversion for artefact reduction, as well as demonstrating the accuracy and robustness of the practical implementation. Producing pre-defined movements that can be compared to those likely to occur in clinical practice can simulate more complex artefacts under experimental conditions. The combination of probe coupling and physiological effects that are expected to occur will result in a level of artefact reduction that not only depends on the system performance, but also the origin of the artefacts. In this way we can separate the

verification of system performance from the investigation of the scope of this technique for clinical application.

4.3.1a Probe-coupling artefact

The generation of pure probe coupling artefact is relatively straightforward when using a standard transmission mode finger probe. Pressure applied to the finger probe clip lifts the opto-electronics away from the finger, therefore simulating probe-coupling artefact. Obviously, care must be taken during the experiments not to allow the probe body to exert any pressure on the finger, which would cause physiological artefacts, or to produce movements that alter the optical field of view. Operating within these constraints, a large range of pure probe coupling artefacts can be generated. Movement distance can be controlled by the clip pressure, whilst varying the average distance between the finger and the opto-electronics simulates the effects of differing source intensity, coupling efficiency and physiological variability.

4.3.1b Complex artefacts

Complex movements can be generated in a multitude of ways in a clinical environment, making it impossible to perform an exhaustive analysis. In addition a large number of movements, such as the raising or lowering of the arm, will only produce physiological artefacts because they do not have a direct effect on the finger probe. The range of physical movements that can directly disturb the probe can be decomposed into a number of fundamental components, which can then be independently tested. A specific complex movement could then be expressed as a combination of these fundamental components. The fundamental components that have been identified for this study are:

- Bending the finger at the distal-phalangic joint
- Bending the finger at the mid-phalangic joint
- Waving, or pseudo-random motions of the hand

Other movements that may appear fundamental, such as bending the finger at the knuckle or exerting pressure on the probe body are not considered here because the artefacts that they produce are predominantly physiological and therefore do not fall within the scope of our definition of artefact in PPG studies. Indeed, in contexts other then pulse oximetry many of these physiological dynamics may be considered valuable information (see section 1.2.2).

4.3.1c Data acquisition

Data sets of 25 seconds in length were utilised for each class of experiment. The subject made pre-defined movements of specific type, magnitude and frequency throughout the experiment. Performing time dependent artefact quantification ensured time invariance across the data set, with highly variable data being discarded. The artefact measurement described previously could then be applied to both the unequalised and equalised signals for a number of data sets, resulting in pairs of values that are indicative of the artefact reduction performance. The difficulty in generating artefact of a pre-defined magnitude dictated that a number of experiments inducing artefacts of varying size were performed, with the resulting data sets being sorted and ranked according to the measured artefact. A plot of these artefact pairs can then demonstrate individual data set performance, as well as indicating trends in the performance with artefact severity.

4.3.2 Experimental results

The level of artefact reduction achieved under conditions of pure probe-coupling artefact is demonstrated in figure 4.11, with artefact levels of up to 70% of signal power being effectively reduced to a level that is below the minimum sensitivity of the artefact measure. There is no obvious trend in the artefact measure of the equalised signals with increasing artefact, suggesting a high degree of equalisation with almost zero residual artefact. This hypothesis can be confirmed by visual inspection of the data sets, which exclusively display an uncorrupted equalised signal. Figure 4.11 also shows that, in all cases, the equalised signal has a lower level of artefact than the unequalised signal.

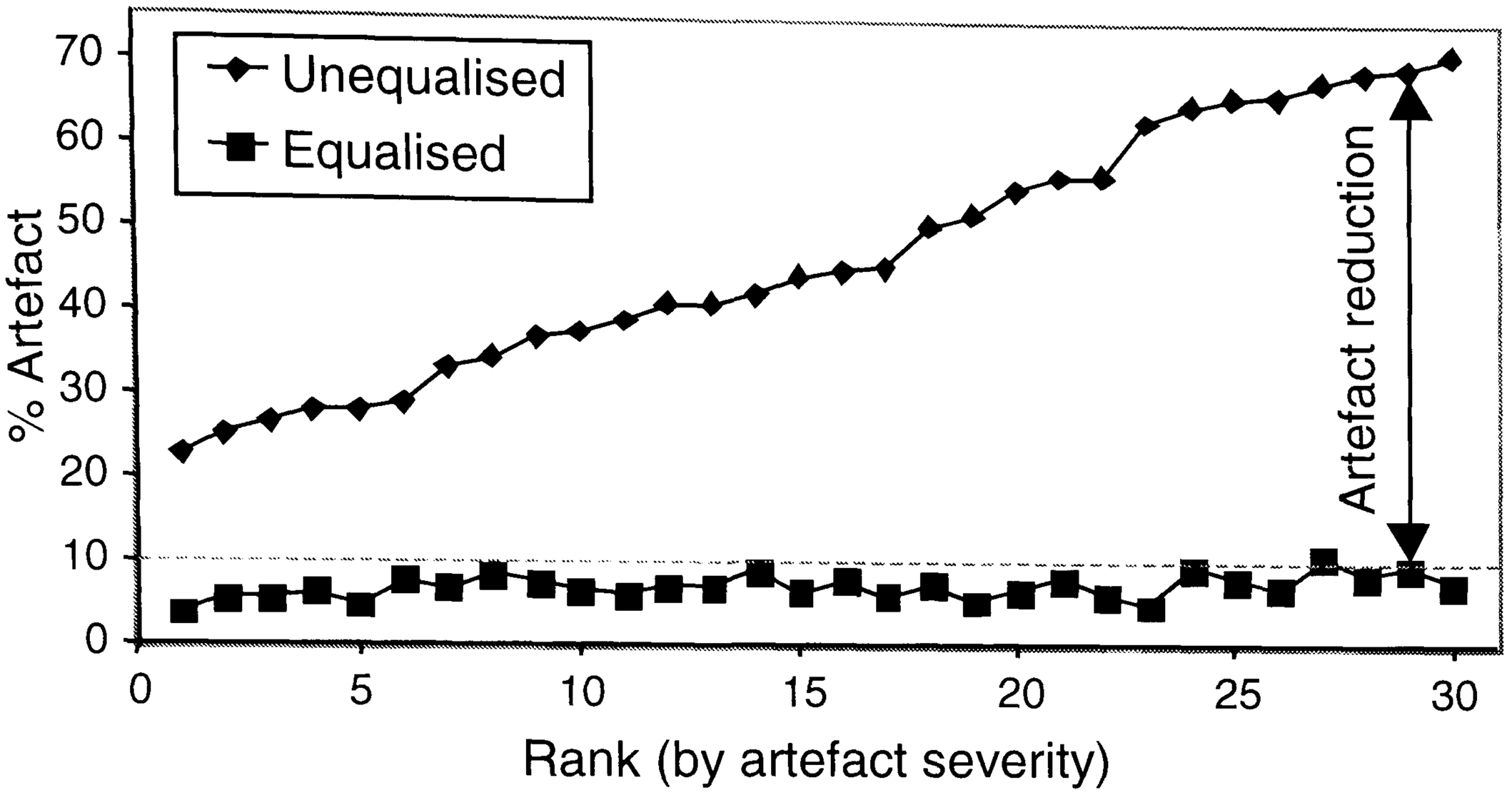


Figure 4.11 – Equalisation of pure probe-coupling artefact

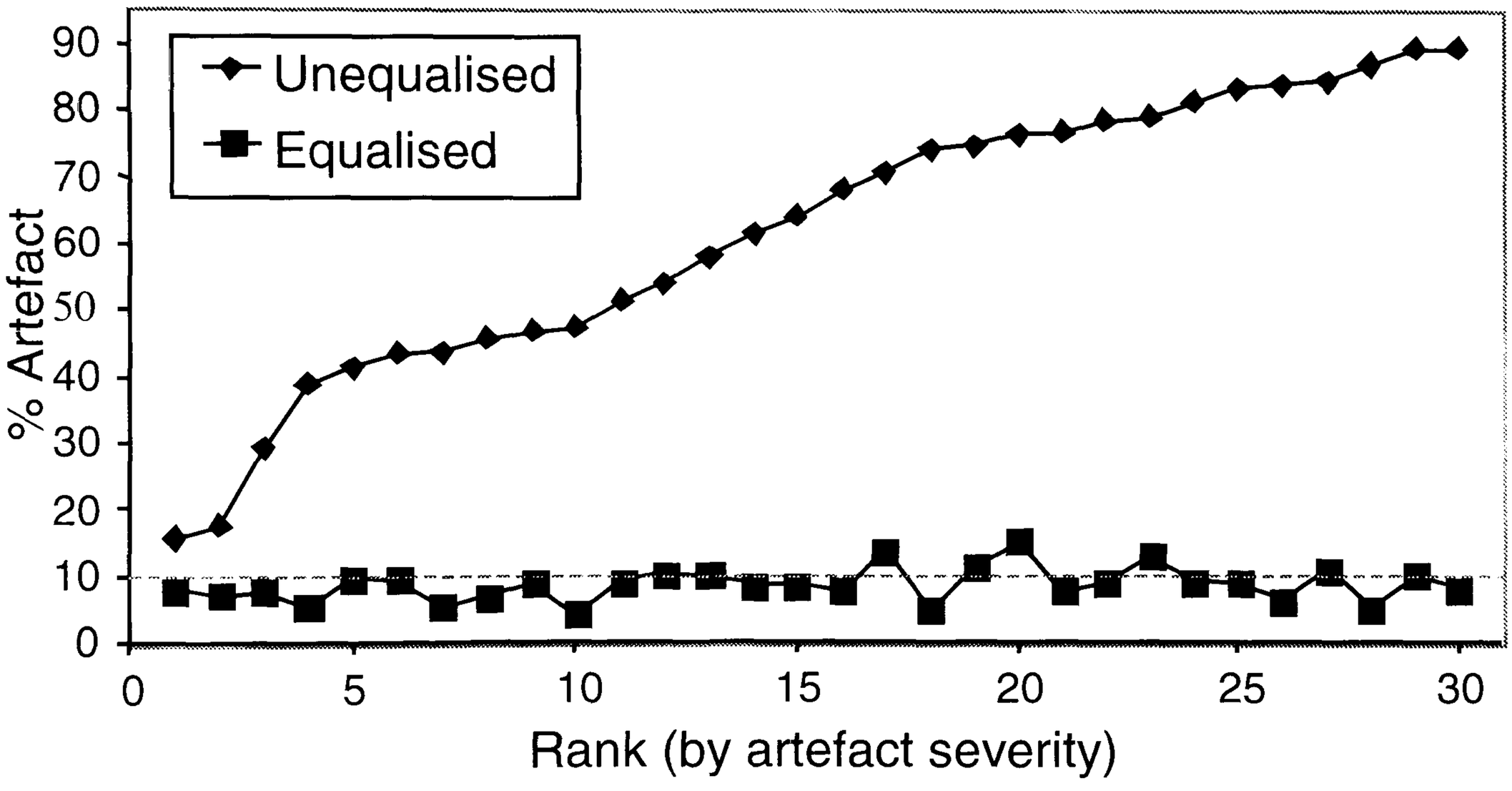


Figure 4.12 – Equalisation of artefact caused by bending the finger at the distal-phalangic joint

4.3.2a Distal-phalangic artefact

Equally impressive levels of artefact reduction were achieved during periods of finger bending at the distal-phalangic joint (figure 4.12), with artefact levels up to 90% of

signal power being reduced to below the sensitivity of the measurement. It should be noted that the increased scale of artefact corruption is indicative of the inherent severity of artefacts generated in this manner. Again, the equalised signal shows no trend with increasing artefact severity and the artefact level was reduced by the non-linear methodology in all cases. This suggests that artefacts generated in this way are predominantly probe coupling in nature and agree well with the assumptions made in the heuristic artefact model.

4.3.2b Mid-phalangic artefact

Bending the finger at the mid-phalangic joint resulted in artefacts that were predominantly removed by the equalisation process, with some measurable residual artefact left on the equalised signal for severe artefacts in excess of about 40% of signal power. This is highlighted by the trend in residual equalised artefact with artefact severity indicated in figure 4.13. This result suggests that whilst this class of movements induces artefacts that are predominantly probe-coupling in nature, more severe movements can cause additional effects, such as physiological dynamics or skewing of the probe geometry, that cannot be removed by this methodology. It is

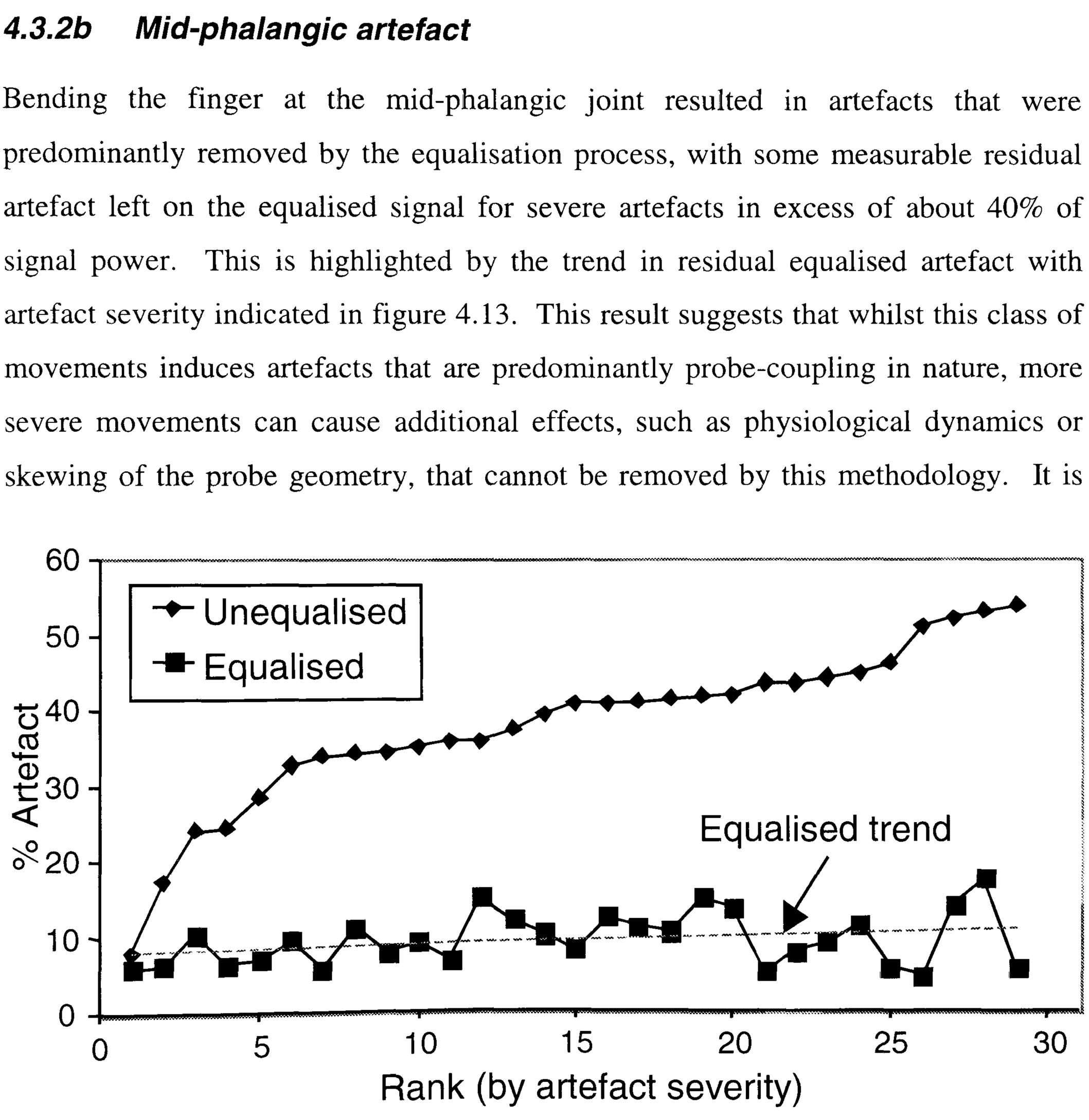


Figure 4.13 - Equalisation of artefact caused by bending the finger at the mid-phalangic joint

also worth noting that artefacts created in this way are not as large as those created by previous methods.

4.3.2c Pseudo-random waving artefact

Waving the hand in a pseudo-random manner resulted in artefact that was reduced in all cases, with almost zero residual artefact when the movement was not severe. Severe artefacts of this class were not completely removed by this technique, again demonstrated by the trend in equalised residual artefact with increasing artefact severity (see figure 4.14). It is likely that the geometrical skewing that can result from rapid and exaggerated hand movements is responsible for this effect. It should be noted, however, that even in cases of incomplete artefact removal there is still a significant improvement in signal quality that results from the removal of all probe-coupling effects.

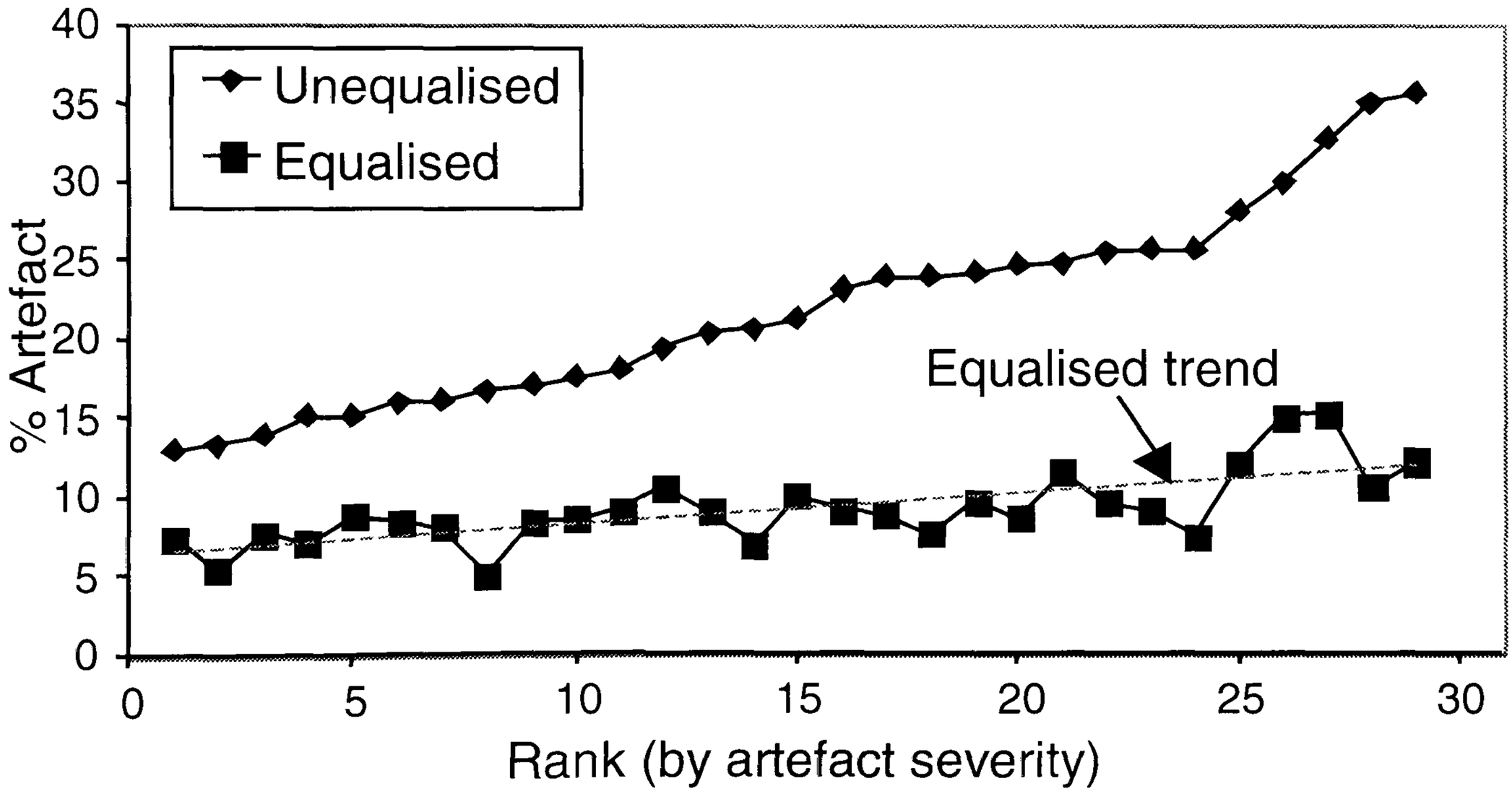


Figure 4.14 – Equalisation of artefact caused by pseudo-random hand movements (waving)

4.3.2d Conclusions

These results demonstrate the high degree of equalisation possible with the non-linear reduction methodology in the context of transmission-mode finger PPG. The fundamental component movements that have been identified as a source of motion artefacts are in no way complete, since they neglect movements that would cause predominantly physiological effects and are therefore not relevant to this study. However, they do represent a major class of motion artefacts encountered in realistic situations and therefore act as a reasonable performance indicator for the practical applicability of the artefact reduction system.

The performance of the artefact reducing methodology should also be highlighted under conditions of very mild artefact. With all classes of artefact, slight artefacts (less than about 20% of signal power) were reduced or removed. In these cases the artefact corruption may not be obvious, in terms of artefact recognition systems, manual observation of signal quality or by interpretation. The fact that artefact reduction occurred in these cases demonstrates the presence of mild signal corruption that could cause inaccuracies in interpretation for applied technologies such as pulse oximetry, even in the absence of end-user perception of artefact problems. This suggests that all PPG technologies would benefit from this artefact-reduction methodology, regardless of whether there is a perceived artefact problem.

4.4 SUMMARY

The practical artefact-reducing PPG system has enabled us to investigate the validity of several key assumptions and theoretical results used to obtain a model inversion. Re-normalisation of the PPG signals with optical information has been practically observed and demonstrations of motion and ambient artefact reduction have been achieved. These results have served to verify the formulations of previous sections.

By definition and application of a quantitative artefact measure to relevant experimental data we have achieved several useful insights into the scope and performance of the non-linear artefact reducing methodology. The near-perfect

removal of a large range of pure probe-coupling artefacts has served to verify the heuristic assumptions about the nature of this class of artefact, as well as demonstrating the success and accuracy of both the model inversion for artefact reduction and the practical implementation utilised. In addition, investigation of the spectra of corrupted PPG signals from controlled experimental conditions has further justified the multiplicative nature of motion artefact. Investigation of fundamental non-physiological artefact producing movements in the context of a transmission-mode finger probe has demonstrated that a large number of artefacts encountered in a clinical environment are likely to be predominantly probe-coupling in nature, and therefore can be effectively removed by this approach. In addition, a significant improvement in signal quality may still be observed when corruption is by artefacts that are only partially of probe-coupling origin. This implies that the non-linear methodology could be used in clinical PPG technologies where artefacts currently cause practical problems. The definition of artefact within the context of non-fundamental PPG technologies means that other physiological signal components could still cause dynamics that would be interpreted as artefact. However, the removal of probe-coupling artefact is a general process that does not affect the additional use of extant artefact reduction solutions to suppress the relatively minor physiological effects.

The success and generality of this artefact reduction method makes it suitable for application to any PPG system, whether fundamental in nature or a derived technology, where subject movement may cause motion artefact that degrades the signal-to-noise ratio of the PPG signals. In addition, the analogue and continuous nature of the technique can result in improvement of signal quality under conditions of very mild artefact that may not be recognised by other techniques or by obvious interpretation errors. The effect of this mild artefact would be to reduce the accuracy of interpretation for technologies such as pulse oximetry, even when gross interpretation of factors such as pulse rate give reasonable accuracy and there is therefore no end-user perception of artefact corruption. This suggests that non-linear artefact reduction may be used to improve the accuracy of conventional PPG techniques, regardless of whether or not artefact corruption has been reported.

5 CONCLUSIONS \ DISCUSSION

The primary aim of this thesis has been to investigate the possibility of reducing artefact corruption of photoplethysmographic signals in real time, using an electronic processing methodology that is based upon inversion of a physical artefact model. A general artefact model has been developed and experimentally justified, leading to a non-linear inversion for deterministic artefact reduction. It has been shown that this methodology is successful in removing or reducing a large number of artefacts encountered in clinically relevant situations. The consequences of this technique for analysis of the photoplethysmographic signals have also been investigated, resulting in a new method for pulse oximetry. It is further suggested that non-linear artefact reduction may be used not only to improve the accuracy and widen the applicability of all photoplethysmographic technologies, but also to empower measurements that are currently not possible.

5.1 CONCLUSIONS

Biomedical monitoring and diagnosis using photoplethysmography is impeded by signal artefact, which reduces the accuracy of derived indices and restricts the applicability of many useful technologies. Although attempts have been made to address the problem within the context of pulse oximetry, no existing technique is suitable for application to fundamental PPG studies. The majority of these approaches utilise a heuristic assumption that artefact corruption is manifested as an additional intensity component, either explicitly or as implied by the underlying methodologies. It has been demonstrated, with the aid of an experimental investigation, that this hypothesis may not be accurate for the majority of motion artefacts. Instead, a multiplicative model, in which the effect of subject movement is to modulate the received intensity, is more physically appropriate. We have termed this class of signal corruption probe-coupling artefact, because of its physical origins.

The model developed in this thesis to describe both the PPG signals and probe-coupling artefact is consistent with the predominantly used contemporary models, the Beer-Lambert law and Diffusion theory. An additional measurement channel can be utilised to invert this model and achieve deterministic artefact reduction, with the solution method requiring a non-linearity. Practical implementation of this inversion can be greatly simplified with the use of logarithmic transduction or signal processing techniques. A significant proportion of signal artefact encountered in a clinical environment will be removed by the resultant non-linear artefact reduction methodology, with consistent improvement in the observed signal quality. The success and generality of this approach make it suitable for application to any PPG system, whether fundamental in nature or a derived technology, where subject movement causes artefact that degrades the signal-to-noise ratio of the PPG signals.

This artefact reduction methodology can be used to develop a new method for pulse oximetry that possesses inherent insensitivity to motion artefact, by utilising an additional source that provides sufficient information to reduce artefact corruption of both measurement sources. One consequence of this approach is that calibration of

the device will no longer be sensitive to the static signal levels, implying that simplifications can be made in the implementation. Careful selection of source wavelengths can also be used to simplify the resultant calibration algorithm.

The artefact reduction methodology developed in this thesis can also be used to isolate the arterial PPG signal in circumstances that would not otherwise be possible, such as remote mode PPG (see section 5.2.1).

5.2 SUGGESTIONS FOR FUTURE WORK

This section outlines the authors suggestions for future work in this area, highlighting both speculative applications empowered by this artefact reduction technology and possible studies that could be used to verify and extend the models developed in this thesis.

5.2.1 Remote PPG

One possible future application of non-linear artefact reduction is the development of non-contact, or remote, PPG. In this mode the subject is physically separated from the opto-electronic components, which operate in reflection mode. Artefact is inherent in any remote situation, because of the small involuntary movements of the subject. Whilst there are other practical problems in remote mode, such as reduced signal-to-noise ratios, it is artefact corruption that is the primary obstacle to the development of a practical system. Remote PPG measurement could be useful in ambulatory studies and in the monitoring of children or neonates, by removing the requirement of physical probe attachment.

5.2.1a Preliminary investigation

In remote mode PPG the possibility of subject motion inducing physiological dynamics is reduced, because there is no pressure on the subject from the probe. It is therefore suggested that the multitude of small involuntary motion artefacts observed in this mode will be predominantly probe-coupling in nature, and therefore well described by the model of section 2.4. However, voluntary subject movements will

be relatively more severe in remote mode, with a small movement completely altering the field of view of a standard reflection mode probe. In order to test this hypothesis, a custom dual-channel reflection mode PPG sensor was connected to the artefact reducing system. The source power was then set to maximum and the probe was physically separated from the finger of the subject by around 3cm of air gap. Whilst the subject attempted to remain motionless, the hand was not clamped or secured in any way. Although adequate signal levels from the source illumination were obtained, involuntary motion artefacts made it impossible to identify valid arterial pulsations. The equalised version of the signal, however, was much improved and could be used to correctly determine the subject's heart rate (see figure 5.1).

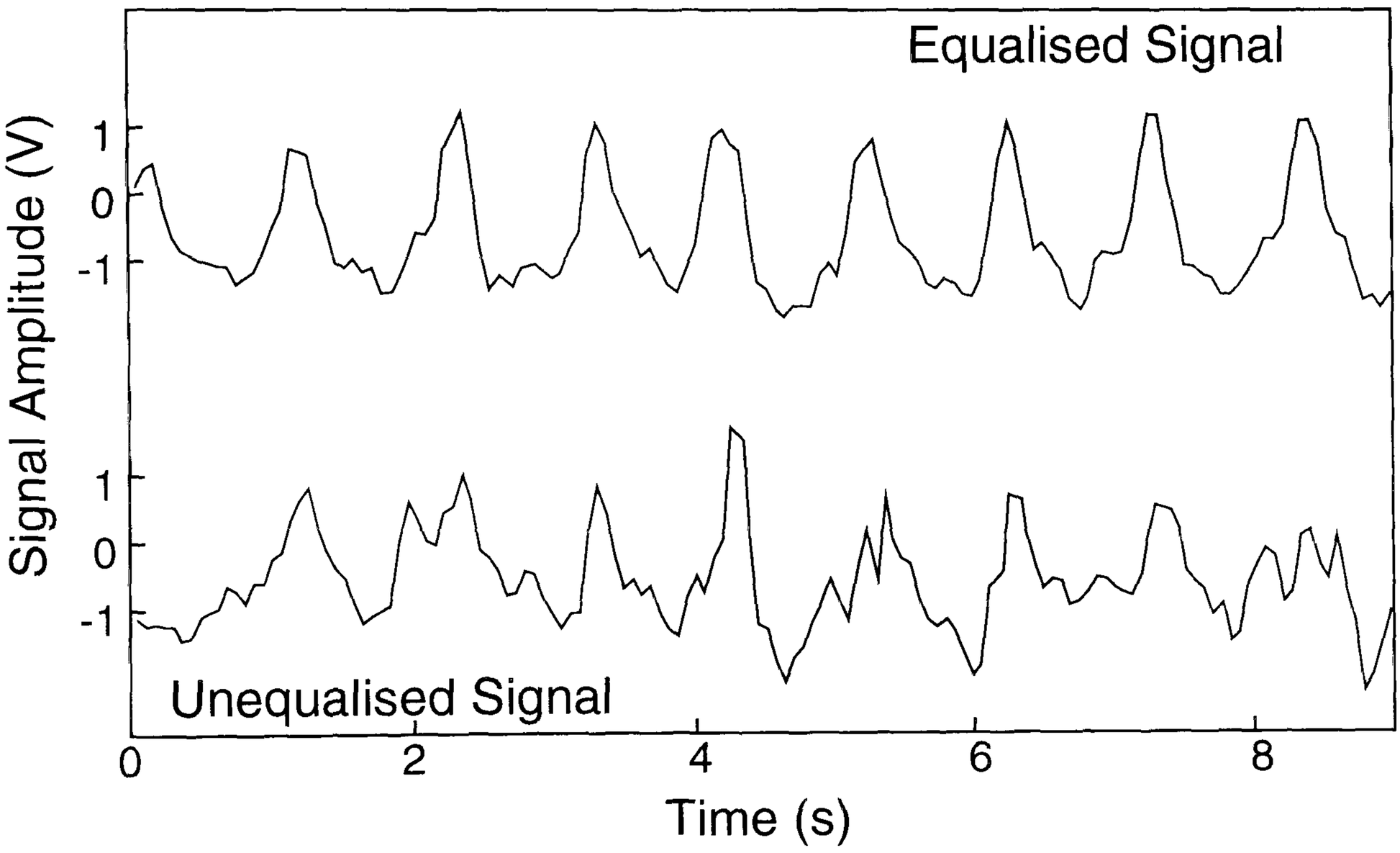


Figure 5.1 – Demonstration of remote mode PPG

Whilst figure 5.1 clearly demonstrates that there is potential in the method, further work would be required to implement a practical system. The primary limiting factor in the performance of this preliminary investigation is the probe design, which was not significantly modified from the conventional reflection mode configuration. It is therefore suggested that a detailed feasibility study of remote mode PPG should be performed before any attempt be made to construct a practical system. This study

should address issues of both probe design and modification to the electronic system to allow for the reduced signal levels and poorer signal-to-noise ratios.

5.2.1b Probe design

There are several fundamental difficulties in the design of a remote PPG probe intended for use in conjunction with the artefact-reduction methodology, namely:

1. Minimising the direct coupling of light between the sources and the receiver, which degrades the artefact reduction performance (see section 3.1.2c).

2. Maximising the probe field of view, so that spatial averaging reduces the observed physiological heterogeneity under conditions of subject movement.

3. Maximising the amount of received light, and therefore the signal-to-noise ratio of the fundamental signals.

Optimal design and evaluation of a probe for remote PPG will therefore require detailed optical modelling. In addition, other factors influencing the success of artefact reduction mentioned previously (section 3.1.2), such as source placement and choice of wavelengths, must also be considered. Careful design and construction of such a probe would ideally be required to perform a quantitative investigation into the feasibility of remote mode PPG.

5.2.2 Extension of the model

Whilst interpretation of the heuristic PPG model presented in section 2.1 could in principle be made from any theoretical perspective desired, the physical applicability of the motion artefact model of section 2.4 will ultimately determine the degree of artefact reduction of any practical implementation. We have already performed a quantitative investigation of the success of artefact reduction using conventional finger probes (section 4.3), concluding that a large number of artefacts encountered in clinical practice are well described by this model. We have also observed that the equalisation process significantly reduces other artefacts that may be less accurately

described by the model. This observation and others may lead to future improvements in the artefact model, by using the existing technology to assist in a re-visited physical investigation (see section 2.4.2).

5.2.2a Additive artefact

The physical investigation of motion artefact performed in this study (section 2.4.2) concluded that it was more physically appropriate to model motion artefact as a modulating factor than it would be to adopt the more common assumption of additive artefact. Whilst this approach has undoubtedly proven successful, the investigation performed was not able to conclusively define the exact form of the artefact corruption, a limitation resulting from the inherent difficulty in the independent measurement and classification of the signal components. However, the artefact reducing PPG system itself empowers a new type of physical artefact investigation; one in which any common multiplicative component of artefact has been removed. This could allow a study of the second-order effects of motion artefact that would otherwise be obscured by the larger modulating effects. One such effect is the possibility of an additive artefact term that is relatively small by comparison with the modulation effects.

5.2.2b Reduction of additive artefact

There are two possibilities for the form of second-order additive artefact effects, namely a wavelength independent additive term or a set of terms that vary with wavelength. The effects of motion artefact reduction on a wavelength-independent additive term have already been considered in section 3.1.3, where we examined the consequences of artefact reduction on residual ambient artefact. It was demonstrated that the equalisation process would reduce any such term by a factor that may be determined from the static signal levels (see equations [3.7] and [3.8]). It would therefore be possible to evaluate the hypothesis of a wavelength-independent additive artefact term by performing a number of artefact-inducing experiments that result in non-ideal artefact reduction. If this hypothesis were appropriate, then intentionally varying the static received signal levels should result in a corresponding variation in

the severity of residual artefact after equalisation. This could be practically achieved by modulating one LED intensity at a fixed (low) frequency and searching for that frequency in the time-dependent artefact quantification (see section 4.2.3b). Of course, careful control of the static signal levels could also be used to ensure effective reduction of additive artefact effects, enabling the experiments of section 4.3 to be repeated and the results quantitatively compared.

5.2.2c Wavelength dependent additive artefact

A similar technique to that used to evaluate the hypothesis of wavelength-independent additive artefact could also be used to examine the possibility of wavelength dependence in a set of additive terms. In this case the LED intensities should be adjusted so that the reduction of additive artefact is poor, and the preceding experiments repeated. Analysis of the time-dependent artefact quantification can then be used to compare the residual artefact after equalisation for a pair of equalised signals that utilise different wavelengths. For example, good correlation between the artefact measures would suggest the possibility of synchronous terms that varied only in magnitude. Note that these experiments would require a three-wavelength implementation that provided a pair of equalised signals, much like that proposed in section 3.2 to create a pulse oximeter with reduced sensitivity to artefact.

5.2.3 Wavelength selection

It has been demonstrated previously (section 3.1.1) that effective isolation of the PPG dynamics following equalisation is dependent upon the use of sources of contrasting tissue coupling characteristics. Using the same notation as before, this requirement can be expressed as

$$\frac{\gamma_j(\lambda)}{\beta_j(\lambda)} \gg \frac{\gamma_k(\lambda)}{\beta_k(\lambda)} \; . \qquad \textbf{[5.1]}$$

Since interpretation of these coupling coefficients can be performed by a multitude of physical models, equation [5.1] suggests that the choice of sources in a practical system may require a non-trivial optimisation.

5.2.3a Beer-Lambert law simplification

It is instructive to examine the source requirements for the case of a simplistic Beer-Lambert law interpretation of the model (see section 2.2.2). In that case we can express the ratio of coupling coefficients in terms of the wavelength-dependent absorbance of blood, and write equation [5.1] as

$$\mu_{blood}(\lambda_j) \rangle\rangle \mu_{blood}(\lambda_k) . \qquad \textbf{[5.2]}$$

Clearly, it is straightforward to satisfy equation [5.2] using knowledge of the absorption spectrum of whole blood. Although this interpretation may give a rough indication of suitable wavelengths, it should certainly not be considered in isolation as it neglects the scattering effects of both blood and tissue.

5.2.3b Diffusion theory interpretation

More insight into suitable choices for the wavelengths can be obtained by referring to the diffusion theory interpretation of the model (section 2.2.3). In contrast with the Beer-Lambert interpretation, which separates transmission through blood and other tissue constituents, this approach models the entire anatomy as a homogenous medium of blood-perfused tissue. In that case the ratio of coupling coefficients was characterised with wavelength dependency with the use of a parameter $\delta_d(\lambda)$, often referred to as the penetration depth for unscattered photons. With reference to the standard diffusion parameters, we can write $\delta_d(\lambda)$ as

$$\delta_d(\lambda) = \sqrt{\frac{3}{\mu_a(\lambda)\left(\mu_a(\lambda) + \mu_s(\lambda)\left(1 - g(\lambda)\right)\right)}} . \qquad \textbf{[5.3]}$$

Equation [5.3] demonstrates that optimal choice of source wavelengths requires knowledge not only of the absorption spectrum of the perfused tissue, but also of the wavelength dependence of the scattering properties. It has been demonstrated previously that both the scattering properties of whole blood[87] and bloodless tissue[83] are wavelength dependent, although no complete spectra have been found for perfused tissue. It is therefore suggested that both further modelling and experimentation is required to identify optimal pairs of source wavelengths for artefact reduction.

5.2.4 Calibrated pulse oximetry

In section 3.2 a new method for pulse oximetry was introduced that possesses inherent insensitivity to motion artefact, achieved by utilising a pair of equalised signals to calibrate and calculate the arterial oxygen saturation. It was also highlighted that this new method represents a generalisation of the conventional formulation, to which it can be demonstrated to simplify under certain conditions (see section 3.2.3). Rigorous evaluation of this new method would require three-wavelength circuitry that could equalise a pair of measurement wavelengths with respect to a third control wavelength (see section 3.2.2d). Although the artefact-reducing PPG system described in section 3.3 is capable of equalising only one measurement channel, the principle of operation is the same for a dual-channel system. The detailed implementation described in appendix I could therefore be extended for dual-channel operation, enabling experimental investigation of this new technique. The following sections suggest possible investigations that could be performed with such a system, resulting in sufficient information to construct a calibrated pulse oximeter with insensitivity to artefact.

5.2.4a Wavelength selection

In section 3.2.2b a ratio of ratios, equivalent to the quantity conventionally used to calculate oxygen saturation, was derived from a pair of measurement signals that had both been equalised with respect to a third control wavelength. Numerical estimation of this value was rendered independent of the tissue characteristics at the control

wavelength with the use of a control source that displays substantially reduced absorption by blood with respect to both measurement wavelengths. It was also demonstrated that more careful selection of wavelengths would enable further simplification and a reduction in the number of additional calibration factors (section 3.2.2c). The requirement for this simplification to be valid (equation [3.15]) can be expressed as

$$\frac{\gamma(\lambda_3)}{\beta(\lambda_3)}\mu_{blood}(\lambda_3) \langle\langle \frac{\gamma(\lambda_2)}{\beta(\lambda_2)}\mu_{blood}(\lambda_2) . \qquad \textbf{[5.4]}$$

Assuming the same physical model for the pulsations as equation [3.10], the requirement of equation [5.4] can be expressed in terms of the dynamics of logarithmically transformed intensities from the relevant wavelengths (see equation [2.28]),

$$\left|v(t,\lambda_3)_{AC}\right| \langle\langle \left|v(t,\lambda_2)_{AC}\right| . \qquad \textbf{[5.5]}$$

Equation [5.5] indicates that a study of the magnitudes of the dynamics of the transformed intensities could be used to evaluate a particular choice of wavelengths. In addition, the dependence of the ratio

$$\mathrm{B} = \frac{\left|v(t,\lambda_3)_{AC}\right|}{\left|v(t,\lambda_2)_{AC}\right|} \qquad \textbf{[5.6]}$$

on independently measured oxygen saturation could also be explored, enabling estimation of the size of any errors introduced by the simplifications.

5.2.4b Additional calibration factors

Having chosen suitable wavelengths for the measurement and control sources, it would then be possible to empirically calibrate the new pulse oximeter using induced hypoxia experiments. The resultant calibration would then define numerical values

for both the conventional and additional calibration parameters. We have assumed that the additional calibration parameters will be substantially independent of any geometric coupling condition or individual tissue type, resulting in a broad applicability of the calibrated measurement. It would be desirable, however, to test this hypothesis by quantifying the variation in these calibration parameters with different individuals or probe positions. Unfortunately there will be significant variation in the conventional calibration parameters between individuals, resulting from a multitude of anatomical and physiological variations that usually serve to restrict the applicability of conventional pulse oximetry (see section 1.3.5 for details). This observation complicates measurement of the variation in the additional parameters, which are coupled to variations in the conventional parameters by the empirical method of calibration.

A better solution is to directly compare the applicability of the conventional and modified formulations. This could be achieved by calibrating a conventional pulse oximeter and an artefact-reducing pulse oximeter using the same subject, maximum degree of induced hypoxia and curve-fitting algorithms. Measurements of calibrated oxygen saturation from a range of subjects could then be performed with both devices, and the degree of accuracy of each device determined by comparison with an independent measurement of the oxygen saturation. The applicability of the calibration schemes could then be evaluated under controlled experimental circumstances.

5.2.4c Insensitivity to artefact

Comparison of the accuracy of conventional and artefact-reducing pulse oximetry under artefact inducing conditions could then be used to quantify the increase in insensitivity to artefact of the new method. Whilst this comparison could in principle be estimated from the quantitative investigation of PPG artefact reduction performed in section 4.3, there are advantages to directly performing the oxygen saturation measurements. Firstly, the fundamental lower limit on the sensitivity of the PPG artefact measurement developed in section 4.2.2 precludes absolute quantification of

the degree of artefact reduction. This restriction would not exist for pulse oximetry measurements, because the oxygen saturation can be measured independently. Secondly, the effect of artefact on the accuracy of conventional pulse oximeter output, which was explored in section 2.4.3b, depends upon the actual oxygen saturation and the calibration parameters of the device. This results in a highly non-linear and unpredictable relationship between PPG artefact corruption and the degree of error in pulse oximeter output. It is therefore more instructive to perform statistical comparisons of the actual error in the output of both devices.

REFERENCES

[1] Hertzman A.B, "Photoelectric plethysmograph of the fingers and toes in man.", Proc. Soc. Exp. Biol. Med., **37**, pp. 529 (1937)

[2] Cartwright C.M, "Infrared transmission of the flesh", J. Opt. Soc. Amer., **20**, pp. 81-84 (1930)

[3] Yelderman M, New W, "Evaluation of pulse oximetry", Anesthesiology, **59**, pp. 349-352 (1983)

[4] Hurch A, Hurch R, Kourg V, "Limitations of pulse oximetry [letter]", Lancet, **1**, pp. 357-359 (1988)

[5] Hertzman A.B, "The blood supply of various skin areas as estimated by the photoelectric plethysmograph", Am. J. Physiol., **124**, pp. 329-340 (1938)

[6] Abramowitz H.B, Queral L.A, Flinn W.R, Nora P.F, "The use of photoplethysmography in the assessment of venous insufficiency: a comparison to venous pressure measurement", Surgery, **86**, pp. 434-441 (1979)

[7] Barnes R.W, "Noninvasive diagnostic techniques in peripheral vascular disease", Am. Heart J., **97**, pp. 2-9 (1979)

[8] Blažek V, Schmitt H.J, Schultz-Ehrenburg U, Kerner J, "Digitale Photoplethysmographie (D-PPG) für die Beinvenendiagnostik", Medizinischtechnische Grundlagen, Phlebol. u., **18**, pp. 91-97 (1989)

[9] Trafford J. de, Lafferty K, "What does photoplethysmography measure? [letter]", Med. Biol. Eng. Comput., **22**, pp. 479-480 (1984)

[10] Challoner A.V.J, "Photoelectric plethysmography for estimating cutaneous blood flow" In: Rolfe P, "Noninvasive physiological measurements", **1**, London Academic, pp. 125-151 (1979)

[11] Lindberg L.G, Öberg P.Å, "Photoplethysmography. II. Influence of light source wavelength", Med. Biol. Eng. Comput., **29**, pp. 48-54 (1991)

[12] Aoyagi T, Kishi M, Yamaguchi K, Wantanabe S, "Improvement of the ear piece oximeter", Abstracts Jpn. Soc. Med. Electronics Biol. Engineer, pp. 90 – 91 (1974)

[13] Steinke J.M, Shepherd A.P, "Role of Light Scattering in Whole Blood Oximetry", IEEE Trans. Biomed. Eng., **33**(3), pp. 294-301, (1986)

[14] Schmitt J.M, "Simple Photon Diffusion Analysis of the Effects of Multiple Scattering on Pulse Oximetry", IEEE Trans. Biomed. Eng., **38**(12), pp. 1194-1203 (1991)

[15] Fronek A, Blažek V, "Automatic calibration of the D-PPG: Its meaning and effectiveness", Report by Third Annual Congress of NASP, Phoenix, USA (1990)

[16] Nitzan M, de Boer H, Turivenko S, Babchenko A, Sapoznikov D, "Spontaneous oscillations in the peripheral circulation system, as measured by photoplethysmography", SPIE, **2328**, pp. 188-195 (1994)

[17] Yoshiya I, Shimada Y, Tanaka K, "Spectrophotometric monitoring of arterial oxygen saturation in the finger tip", Med. Biol. Eng. Comput., **18**, pp. 27-32 (1980)

[18] Nakajima K, Tamura T, Miike H, "Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique", Med. Eng. Phys., **18**(5), pp. 365-372 (1996)

[19] Yamakoshi K, Shimazu H, Shibata M, "A new oscillometric method for indirect measurement of systolic and mean arterial pressure in the human finger", Med. Biol. Eng. Comput., **20**, pp. 307-313 (1982)

[20] Shimazu H, Fukuoka M, Ito H, "Noninvasive measurement of beat-to-beat vascular visccoelastic properties in human fingers and forearms", Med. Biol. Eng. Comput., **23**, pp. 43-47 (1985)

[21] Schmeink U, Schmeink Th, Lossau I, Roschansky V, "D-PPG and Duplex-ultrasound: Correlation of results in the diagnosis of deep vein insufficiency", in *Proceedings of the Seventh International Symposium CNVD (Computer-aided Noninvasive Vascular Diagnostics)*, V. Blažek and U. Schultz-Ehrenburg Eds. (Verein Deutscher Ingenieure, Dűsseldorf, Germany, 1998) **263**, pp 161-164 (1997)

[22] Blažek V, Schultz-Ehrenburg U, "Quantitative Photoplethysmography Basic facts and examination tests for evaluating peripheral vascular functions", VDI (Verein Deutscher Ingenieure) Verlag, **20**(192) (1996)

[23] Millikan GA, "Oximeter, instrument for measuring continuously oxygen saturation of arterial blood in man", Rev. Sci. Instrum., **13**, pp. 434 (1942)

[24] Zijlstra W.G, Buursma A, Meeuwsen-van der Roest P, "Absorption spectra of Human Foetal and Adult Oxyhemoglobin, De-Oxyhemoglobin, Carboxyhemoglobin, and Metahemoglobin", Clin. Chem., **37**(9), pp. 1633-1638 (1991)

[25] Rodrigo F.A, "The determination of the oxygenation of blood in-vitro by using reflected light", Am. Heart J., **45**, pp. 809-822 (1953)

[26] Manheimer P.D, Casciani J.R, Fein M.E, Nierlich S.L, "Wavelength selection for low-saturation pulse oximetry", IEEE Trans. Biomed. Eng, **44**(3), pp. 148-158 (1997)

[27] Takatani S, Davies C, Sakakibara N, Zurik A, "Experimental and clinical evaluation of a noninvasive reflectance pulse oximeter sensor", J. Clin. Monit., **8**(4), pp.257-266 (1992)

[28] Polyanyi M.L, Hehir R.M, "New reflectance oximeter", Rev. Sci. Instrum., **31**, pp. 401-403 (1960)

[29] Aoyagi T, "Apparatus for measuring predetermined data of living tissue", United States Patent No:5385143 (1995)

[30] Tremper K.K, Barker S.J, "Pulse oximetry", Anesthesiology, **70**, pp. 98-108 (1989)

[31] Runciman W.B, Webb R.K, Barker L, Currie M, "The Pulse Oximeter: Applications and Limitations – An Analysis of 2000 Incident Reports", Anaesth. Intensive Care, **21**, pp. 543-550 (1993)

[32] Taylor M.B, Whitwam J.G, "The accuracy of pulse oximeters", Anaesthesia, **43**, pp. 229-232 (1988)

[33] Shimada Y, Yoshiya I, Oka N, Mamaguri K, "Effects of multiple scattering and peripheral circulation on arterial oxygen saturation, measured with a pulse-type oximeter", Med. Biol. Eng. Comput., **22**, pp. 475-478, (1984)

[34] Kramer K, Elam J.O, Saxton G.A, Elam W.N, "Influence of oxygen saturation, erythrocyte concentration and optical depth upon the red and near-infrared light transmittance of whole blood", Am. J. Physiol., **165**, pp. 229-246, (1951)

[35] Grace R.F, "Pulse Oximetry - Gold standard or false sense of security?", Med. J. Aust. , **160**, pp. 638-644 (1994)

[36] Trivedi N.S, Ghouri A.F, Lai E, Shah N.K, Barker S.J, "Pulse oximeter performance during desaturation and resaturation: A comparison of seven models.", J. Clin. Anesth. , **9**(3), pp. 184-188 (1997)

[37] Potratz, R.S, "Condensed oximeter system with noise reduction software", United States Patent No:5351685 (1994)

[38] Prosser S.J, "Method and apparatus for improving the accuracy of pulse transmittance oximeter", United States Patent No:5285782 (1994)

[39] De Kock J.P, Reynolds H.J, Tarassenko J, Moyle J.T.B, "The effect of varying LED intensity on pulse oximeter accuracy", J. Med. Eng. Technol., **15**(3), pp. 111-116 (1991)

[40] Severinghaus J.W, Koh S.O, "Effects of anemia on pulse oximeter accuracy", J. Clin. Monit., **6**, pp. 85-88 (1990)

[41] Poets C.F, Southall D.P, "Noninvasive Monitoring of Oxygenation in Infants and Children: Practical Considerations and Areas of Concern", Pediatrics, **93**(5), pp. 737-746 (1994)

[42] Kelleher J.F, "Pulse oximetry", J. Clin. Monit., **5**, pp. 37-62 (1989)

[43] Ralston A.C, Webb R.K, Runciman W.B, "Potential errors in pulse oximeter evaluation. Part II: Effects of changes in saturation and signal quality", Anaesthesia, **46**, pp. 202-206 (1991)

[44] Miller S.E, Kaminow I.P. "Optical Fibre Telecommunications" Vol. II, Academic Press, New York, pp. 487–488 (1988)

[45] Payne J.P, Seveninghaus J.W, "Pulse oximetry", Dorchester: Springer-Verlag, **90** (1986)

[46] Mendelson Y, Kent J.C, "Variations in Optical Absorption Spectra of Adult and Foetal Hemoglobins and Its Effect on Pulse Oximetry", IEEE Trans. Biomed. Eng., **36**(8), pp. 844-848 (1989)

[47] Ralston A.C, Webb R.K, Runciman W.B, ". Potential errors in pulse oximeter evaluation. Part I: Pulse oximeter evaluation", Anaesthesia, **46**, pp. 207-212 (1991)

[48] Severinghaus J.W, Kelleher J.F, "Recent developments in pulse oximetry", Anesthesiology, **76**, pp. 1018-1038 (1992)

[49] Costarino A.T, Davis D.A, Keon T.P, "Falsely normal readings with the pulse oximeter", Anesthesiology, **67**, pp. 830-831 (1987)

[50] Trivedi N.S, Ghouri A.F, Shah N.K, Lai E, Barker S.J, "Effects of motion, ambient light, and hypoperfusion on pulse oximeter function", J. Clin. Anesth., **9**(3), pp. 179-183 (1997)

[51] Block F.E, "Interference in a pulse oximeter from a fibre-optic light source", J. Clin. Monit., **3**, pp. 210-211 (1987)

[52] Norton L.H, Squires B, Craig N.P, McLeay G, McGrath P, Norton K.I, "Accuracy of pulse oximetry during exercise stress testing", Int. J. Sports Med., **13**, pp. 523-527 (1992)

[53] Hayes M.J, Smith P.R,, Barnett D.M, Morgan M.D.L, Singh S, Vara D.D, "Quantitative Investigation of Artefact in Photoplethysmography and Pulse Oximetry for Respiratory Exercise Testing", in *Proceedings of the Seventh International Symposium CNVD (Computer-aided Noninvasive Vascular Diagnostics)*, V. Blažek and U. Schultz-Ehrenburg Eds. (Verein Deutscher Ingenieure, Dűsseldorf, Germany, 1998) **263**, pp. 117-124 (1997)

[54] Poets C.F, Stebbens V.A, "Detection of movement artifact in recorded pulse oximeter saturation", Eur. J. Pediatr., **156**(10), pp. 808-811 (1997)

[55] Wiklund L, Hok B, Stahl K, Jordeby-Jonsson A, "Postanaesthesia monitoring revisited: Frequency of true and false alarms from different monitoring devices", J. Clin. Anesth., **6**(3), pp. 182-188 (1994)

[56] Gaskin L, Thomas J, "Pulse oximetry and exercise", Physiotherapy, Vol. 81, **5**, pp. 254– 61 (1995)

[57] Langton J.A, Hanning C.D, "Effect of motion artefact on pulse oximeters: Evaluation of four instruments and finger probes", Br. J. Appl. Physiol., **65**, pp. 564–570 (1992)

[58] Benoit H, Costes F, Feasson L, Lacour J.R, Roche F, Denis C, Geyssant A, Barthelemy J.C, "Accuracy of pulse oximetry during intense exercise under severe hypoxic conditions", Eur. J. Appl. Physiol., **76**(3), pp. 260-263 (1997)

[59] Hanowell L, Eisele J.H, Downs D, "Ambient light affects pulse oximeters" [letter], Anesthesiology, **67**, pp. 864–865 (1987)

[60] Brodie T.D, Paulus D.A, Winkle W.E, "Infra-red heat lamps interfere with pulse oximeters [letter]", Anesthesiology, **61**, pp. 630 (1984)

[61] Abbott M.A, "Monitoring oxygen saturation in the early recovery phase of general anaesthesia" In: Payne J.P, Severinghaus J.M, "Pulse oximetry" Dorchester: Springer-Verlag, pp. 165-172 (1986)

[62] Malinouska D, "Apparatus and method for use in pulse oximeters", United States Patent No:4807630 (1989)

[63] Swedlow D.B, "Apparatus for the detection of motion transients", United States Patent No:5226417 (1993)

[64] Matthews G. R., "Pulse Responsive Device", International Patent Application WO 91/18550 (1991)

[65] Diab, Mohamed K, "Signal Processing Apparatus", International Patent Application WO 96/12435 (1996)

[66] Hall P.R, "Motion artefact rejection system for pulse oximeters", United States Patent No:4955379 (1990)

[67] Swedlow D. B, "Oximeter with Motion Detection for Alarm Modification", International Patent Application WO 94/22360 (1994)

[68] Visram A.R, Jones R.D. M, Irwin M.G, Bacon-Shone J., "Use of two oximeters to investigate a method of movement artefact rejection using photoplethysmographic signals.", Br. J. Anaesth., **72**, pp. 388-392 (1994)

[69] Hamaguri K, Sakai T, "Multi-wavelength oximeter having a means for disregarding a poor signal", United States Patent No:4714341 (1987)

[70] Stone R.T, "Method and apparatus for calculating arterial oxygen saturation based plethysmographs including transients", United States Patent No:5078136 (1992)

[71] Parker D, "Optical Monitor (Oximeter etc.) with Motion Artefact Suppression", International Patent Application WO 94/03102 (1994)

[72] Yorkey T.J, "Method and Apparatus for Removing Motion Artefact and Noise from Pulse Oximetry", International Patent Application WO 97/00041 (1997)

[73] Proakis J.G, Manolakis D.G, "Digital Signal Processing – Principles, Algorithms, and Applications", Chapter 11, 2nd Edition, Macmillan (1992)

[74] Dumas C, Wahr J.A, Tremper K.K, "Clinical evaluation of a prototype motion artifact resistant pulse oximeter in the recovery room", Anesth. Analg., **83**(2), pp. 269-272 (1996)

[75] Barker S.J, Shah N.K, "The effects of motion on the performance of pulse oximeters in volunteers", Anesthesiology, **86**(1), pp. 101-108 (1997)

[76] Plummer J.L, Ilsley A.H, Fronsko R.R.L, Owen H, "Identification of movement artefact by the Nellcor N-200 and N-3000 pulse oximeters", J. Clin. Monit., **13**(2), pp. 109-113 (1997)

[77] Ishimaru A, "Diffusion of light in turbid material", Appl. Opt., **28**(12), pp. 2210-2215 (1989)

[78] Graaff R, Koelink M.H, de Mul F.F.M, Zijlstra G, Dassel A.C.M, Aarnoudse J.G, "Condensed Monte Carlo simulations for the description of light transport", Appl. Opt., **32**(4), pp. 426-434 (1993)

[79] Star W.M, "Comparing the P3-approximation with diffusion theory and with Monte Carlo calculations of light propagation in a slab geometry", SPIE, vol. IS5, Dosimetry of Laser Radiation in Medicine and Biology, pp. 146-154 (1989)

[80] Morse P.M, Feshbach H, "Methods of Theoretical Physics", McGraw-Hill (1953)

[81] Johnson C.C, "Optical Diffusion in Blood", IEEE Trans. Biomed. Eng., **17**(2), pp. 129-133 (1970)

[82] Van Germert M.J.C, Jacques S.L, Sterenborg H.J.C.M, Star W.M, "Skin Optics", IEEE Trans. Biomed. Eng., **36**(12), pp. 1146-1154 (1989)

[83] Cheong W.F, Prahl S.A, Welch A.J, "A Review of the Optical Properties of Biological Tissues", IEEE J. Quantum Elec., **26**(12), pp. 2166-2185 (1990)

[84] Patterson M.S, Chance B, Wilson B.C, "Time resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties", Appl. Opt., **28**(12), pp. 2331-2336 (1989)

[85] Hopcraft K.I, Smith P.R, "An Introduction to Electromagnetic Inverse Scattering", Kluwer Academic Publishers (1992)

[86] Marble D. R, Burns D. H, Cheung P. W, "Diffusion-based model of pulse oximetry: *in vitro* and *in vivo* comparisons", Appl. Opt., **33**, pp. 1279-1285 (1994)

[87] Sardar D. K, Levy L. B, "Optical Properties of Whole Blood", Lasers Med. Sci., **13**, pp. 106-111 (1998)

APPENDIX

I. ELECTRONIC CIRCUITRY

This appendix describes in detail the design and implementation of the artefact-reducing PPG system introduced in section 3.3. Whilst this description is not exhaustive, it demonstrates both design decisions and methodology important for the overall system performance, as well as providing a technical summary of the circuitry for reference purposes. Complete technical drawings and code listings for programmable devices supply enough information for such a system to be independently reproduced.

I.1 FRONT-END CIRCUITRY

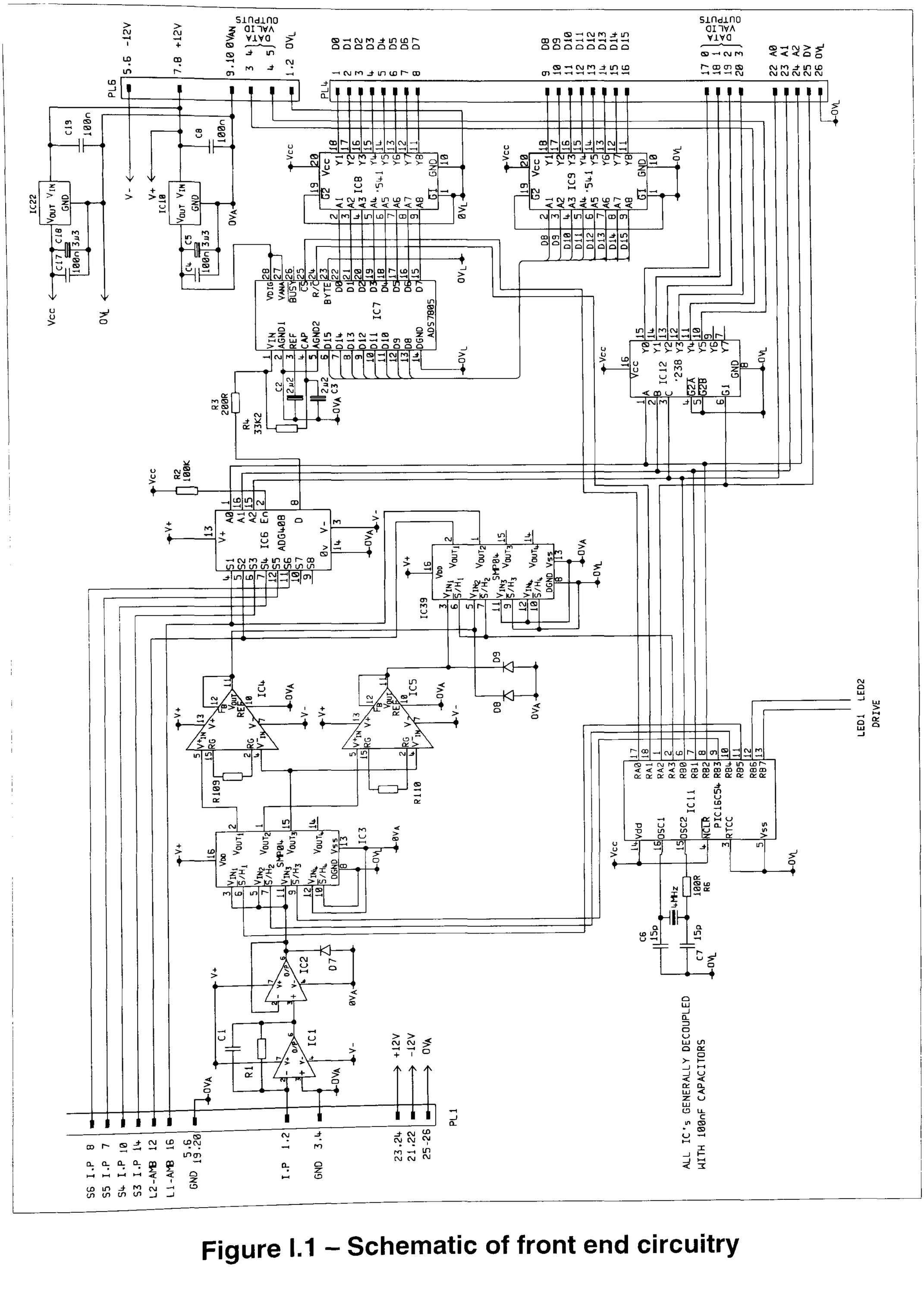


Figure I.1 – Schematic of front end circuitry

This section describes the design and operation of the front-end circuitry used for the dual-channel PPG device. The front-end collects information from the PiN diode receiver, splits it into channels due to the alternating illumination and subtracts any measured ambient light from the desired signals. These signals are then digitally sampled, along with other signals returning from signal-processing stages, to provide output to an external processor. In addition, the front-end circuitry controls the timing of the complete system, providing digital signals to control the LED illumination.

I.1.1 Optical receiver

A PiN diode has a voltage-current characteristic which can be approximated to that of a standard *pn* semiconductor junction with an additional reverse current, I_{PH}, which is known as the photo-current and is proportional to the intensity of light incident on the junction,

$$I_D \approx I_S \left\{ e^{\frac{qV_D}{KT}} - 1 \right\} - I_{PH}, \qquad \textbf{[I.1]}$$

where I_S is the reverse saturation current, K is Boltzman's constant (1.38×10^{-23}), T is the absolute temperature (K) and q is the charge of an electron (1.62×10^{-19} C).

I.1.1a Zero-bias configuration

Equation [I.1] indicates that an open-circuit PiN diode ($I_D = 0$) will generate a voltage, discernible by high-impedance measurement techniques, which is logarithmically related to the photo-current,

$$V_D^{I_D=0} \approx \frac{KT}{q} \ln\left\{ \frac{I_{PH}}{I_S} + 1 \right\}. \qquad \textbf{[I.2]}$$

Whilst the open-circuit voltage of equation [I.2] can be utilised to produce an optical receiver with a logarithmic intensity-to-voltage characteristic, the use of such zero-bias techniques incur a heavy penalty in terms of the achievable bandwidth. This is because they rely on the diffusion of minority carriers within the semiconductor junction to maintain voltage equilibrium, restricting realisable bandwidths to less than 1kHz in most devices. In addition, semiconductor theory predicts the existence of several regions across the device characteristic that would scale the over-simplistic relation of equation [I.2] by an amount dependent on the average static light intensity. This unpredictable nature of the intensity-to-voltage characteristic would therefore make this configuration unsuitable for the present application.

I.1.1b Linear-bias configuration

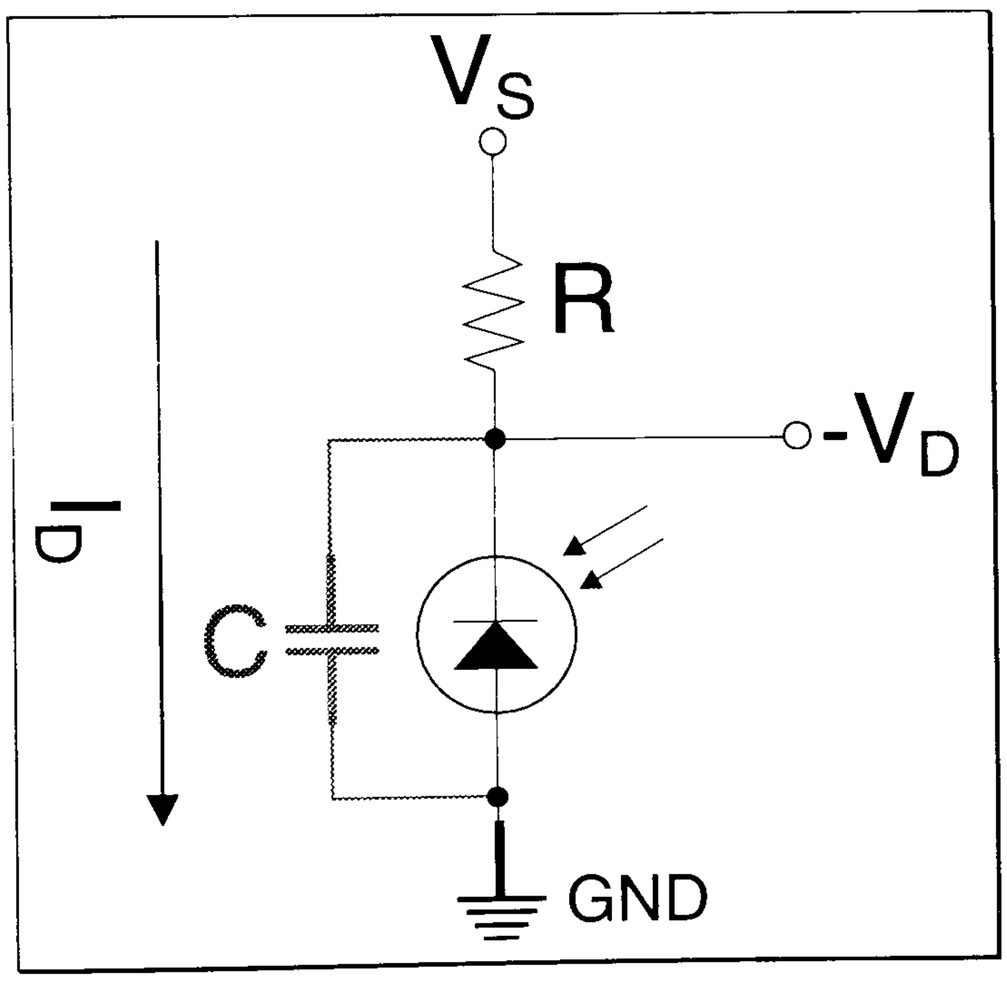


Figure I.2 – Linear bias

Higher bandwidths and linear intensity-to-voltage characteristics can be obtained by biasing the PiN diode, as demonstrated in figure I.2. Application of Kirchoffs voltage law then imposes a linear load-line on the diode characteristic,

$$I_D = \frac{-V_S - V_D}{R}. \qquad \textbf{[I.3]}$$

Simultaneous solution of equation [I.3] with equation [I.1] can then be used to estimate the output voltage, $-V_D$. The graphical solution demonstrated in figure I.3 highlights that if $I_{PH} \rangle\rangle I_S$, as is usually the case, then we may approximate the PiN diode characteristic in the reverse-bias region as simply $-I_{PH}$, enabling us to write the diode voltage as

$$V_D \approx I_{PH} R - V_S. \qquad \textbf{[I.4]}$$

Equation [I.4] demonstrates that the output of such a circuit is linearly related to the incident intensity, with the constant of proportionality, or intensity-to-voltage gain, being determined by the resistance R. Higher values of R therefore result in greater optical sensitivity and an improved signal-to-noise ratio.

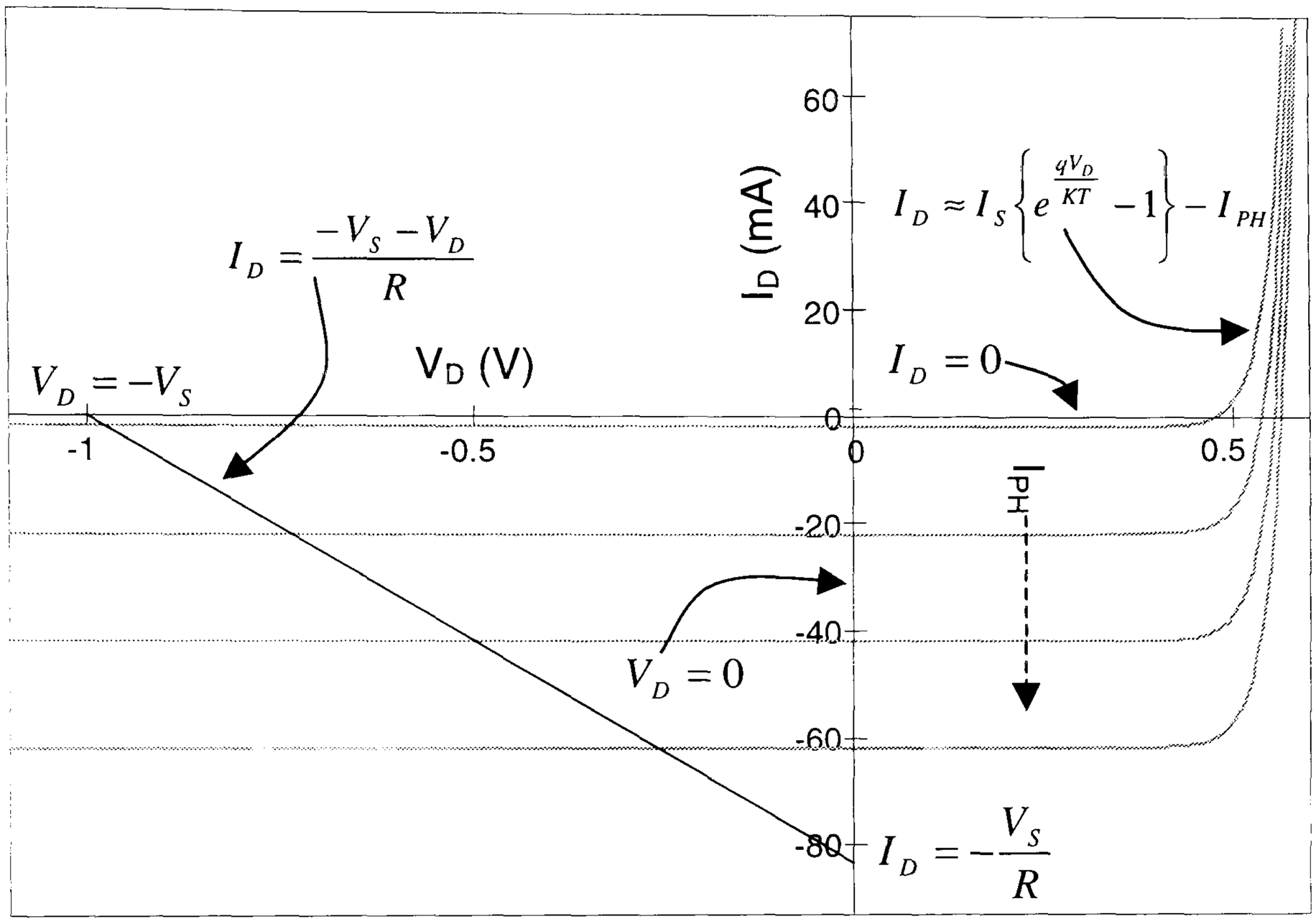


Figure I.3 – PiN diode characteristic with example load-lines

Unfortunately, the linear-bias circuit has a bandwidth restriction that is imposed by a time constant formed between the load resistor, R, and the parasitic capacitance of the PiN diode, C (see figure I.2). Although the value of C in principle decreases with increased bias voltage, V_S, the time constant will be dominated by the load resistor for highly sensitive circuits. Practical realisations of the linear-bias circuit are therefore commonly classified as low-impedance (small R) or high-impedance (large R) implementations. Low-impedance circuits have a high bandwidth, but low sensitivity and signal-to-noise ratio, whilst high-impedance circuits display high sensitivity and signal-to-noise ratio at the expense of reduced bandwidth. Whilst this configuration could in principle be used for the present application, the dual requirements for high speed and high sensitivity could not both be simultaneously fulfilled, resulting in a less than ideal compromise.

I.1.1c *Transimpedance configuration*

A practical circuit that can simultaneously achieve high sensitivity and high bandwidth is the transimpedance configuration, shown in figure I.4.

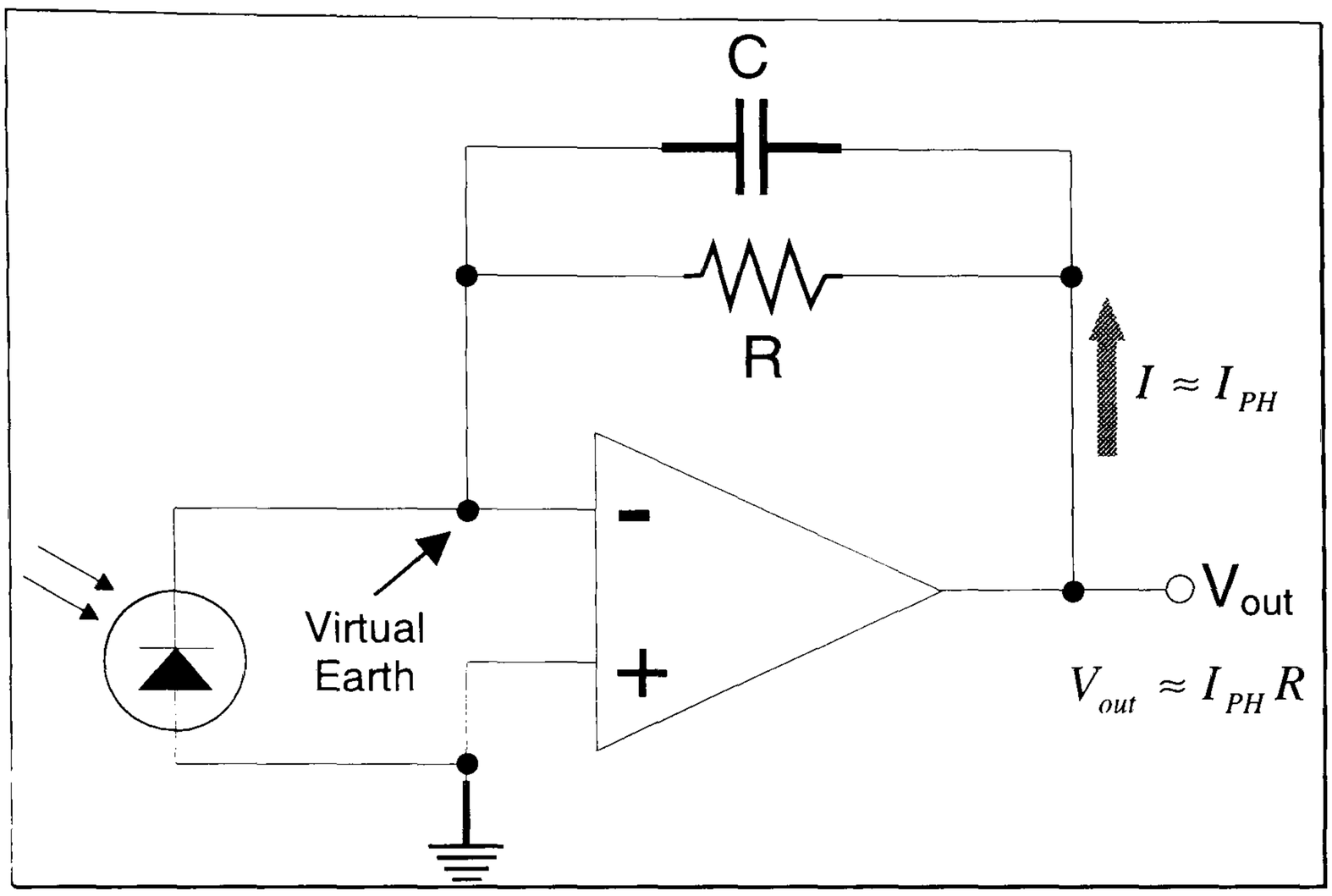


Figure I.4 - Transimpedance configuration

The negative feedback action of the operational amplifier (op-amp) attempts to keep the two inputs at the same voltage, therefore maintaining zero volts across the PiN diode and minimising the effects of any parasitic capacitance. Imposing the condition $V_D = 0$ on the PiN diode characteristic of equation [I.1] results in a diode current that is exactly the inverse of the photo-current, $I_D = -I_{PH}$ (also see figure I.3). Since negligible current flows into the op-amp input terminals, the diode current must flow through the resistor R, resulting in an output voltage that is proportional to the photo-current,

$$V_{OUT} \approx I_{PH} R . \qquad \textbf{[I.5]}$$

A capacitor, C, is included to counter the phase-lag introduced by the PiN diode parasitic capacitance and maintain stability of the feedback loop at all frequencies. This prevents *ringing* of the circuit in response to a step-input (for example as caused by the multiplexing of sources) by increasing the feedback loop phase-margin. This

capacitor may also be used to reduce the bandwidth of the circuit if necessary. It is also possible with many practical devices to increase the bandwidth of the circuit by applying a negative bias voltage to the PiN diode anode, therefore reducing the parasitic capacitance (not shown in figure I.4). Although there is now a finite voltage across the diode (the bias voltage), the feedback action still attempts to minimise any changes in that voltage and therefore continues to minimise the effect of the parasitic capacitance. A bias voltage was not used in this application because the optimal bias voltage and resulting performance are highly dependent upon the specific PiN diode used, and the use of such techniques would reduce the flexibility of the circuit to drive a wide range of probes. The major advantage of this transimpedance configuration for the present application is that a relatively high value of R, and therefore a high sensitivity, can be achieved with a relatively small reduction in bandwidth.

I.1.1d Implementation

Implementation of a transimpedance configuration was achieved using an AD820 op-amp (IC1), which has an extremely low input-bias current and is therefore inherently suitable for PiN diode applications. A unity-gain voltage-follower, also constructed from an AD820 (IC2), is then used to buffer the circuit and de-couple the receiver performance from the loading of subsequent stages. The output of this follower is clamped to ground using a Shottky diode (D7), which prevents the possibility of transient negative excursions that could cause damage to the following sample-and-hold amplifiers.

I.1.2 Circuit timing

Digital timing signals, used to control everything from the source multiplexing, analogue demultiplexing and digital sampling and output, are generated by an 8-bit micro-controller, a PIC16C54 (IC11). This one-time-programmable (OTP) device is available in an 18-pin surface-mount package (SOIC); similar to those used for the packaging of standard logic devices. The major advantage of this methodology over more conventional fixed or programmed logic is the ability of the micro-controller to

generate highly complex timing sequences, with a high degree of timing stability. The PIC16C54 operates from a 4MHz crystal oscillator, resulting in a highly predictable 1μs cycle time. In addition, the flexibility afforded by the use of this approach enables rapid and straightforward modification of the detailed timing of the circuit without the need to physically modify the circuit layout or component choice.

I.1.2a Multiplexing and demultiplexing

Alternate source illumination is used to approximate simultaneous illumination of the vascular bed by both sources. In addition, a time period during which neither source is illuminated is utilised for the measurement of ambient lighting conditions. This results in three basic multiplexing periods, with three corresponding analogue signals following the demultiplexer circuitry. Demultiplexing is accomplished with sample-and-hold circuitry (IC3), thus removing the requirement for low-pass filters that would be needed to smooth the outputs (perform the *hold* function) if conventional analogue switches were used. The common problems of printed-circuit-board (PCB) grounding layout and external capacitor matching that are usually associated with sample-and-hold circuitry have been alleviated with the use of SMP-04 integrated-circuits (ICs). These devices contain four independent matched sample-and-hold circuits, including matched hold capacitors and simplified external grounding.

The detailed timing of the multiplexing / demultiplexing scheme is shown in figure I.5, for a cycle time of 198μs or a repetition frequency of 5.05kHz. Signals **LED1** and **LED2** control the LED driver circuits and are asserted positively to illuminate the relevant LED. The signals $\overline{\textbf{L1 sample}}$, $\overline{\textbf{L2 sample}}$ and $\overline{\textbf{Amb sample}}$ place each of the three sample-and-hold amplifiers in either sample or hold modes, with negative assertion causing the relevant amplifier to sample the incoming signal.

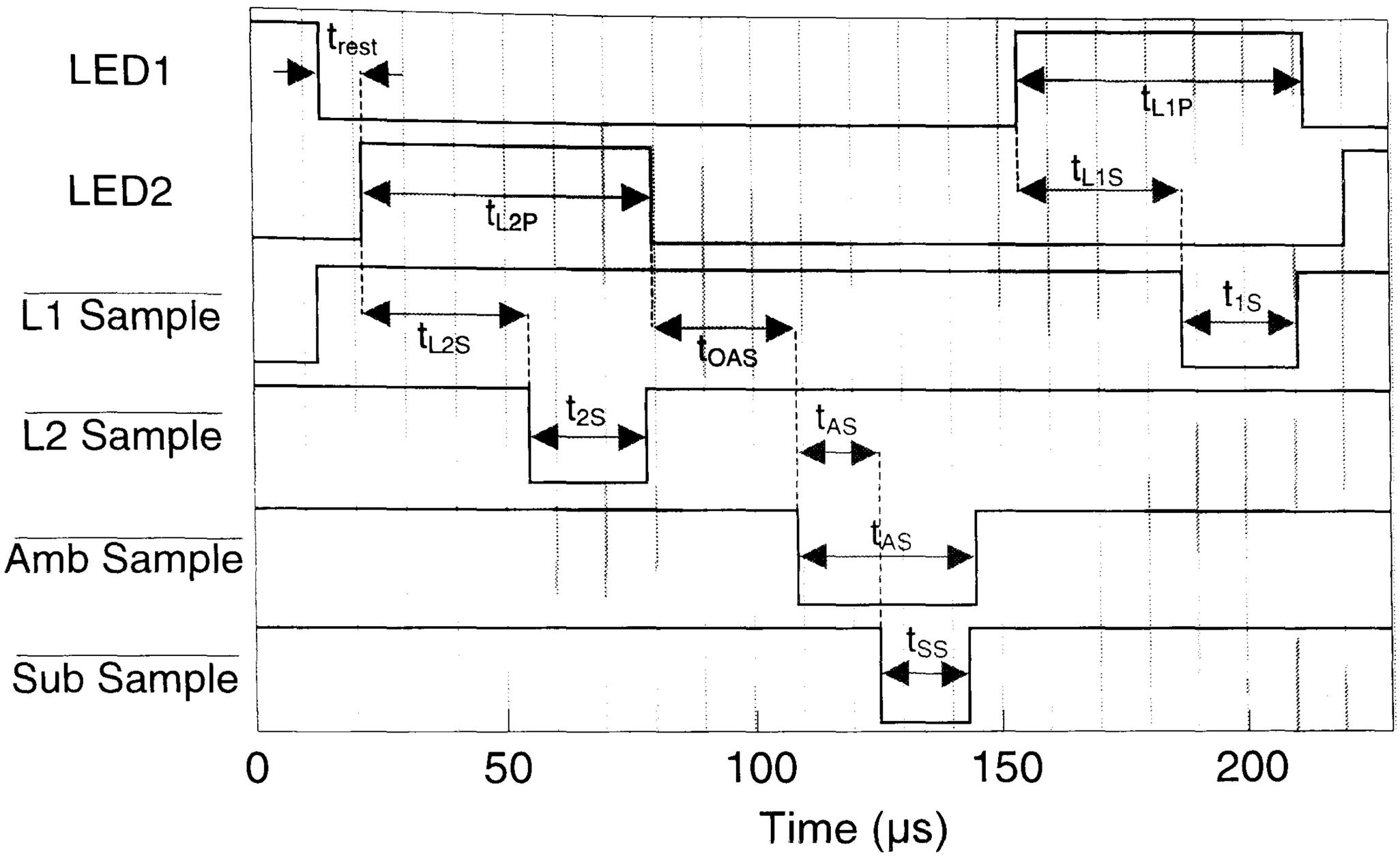


Figure I.5 – Multiplexed circuit timing

Instrumentation amplifiers perform subtraction of the demultiplexed ambient signal from both measurement signals (IC4&5), again with Schottky diodes (D8 & 9) clamping the outputs to ground and preventing the possibility of negative transients. The use of instrumentation amplifiers enables the DC signal levels to be more accurately preserved than would otherwise be possible using op-amp based differencing circuits. In addition, the INA111 amplifiers used are laser-trimmed and therefore provide an inherent matching between the two subtraction stages. More sample-and-hold amplifiers (IC39) then gate the resulting ambient subtracted signals to remove components of the continuous signals that are not valid and allow some settling time for the subtractions to take place. The **$\overline{\text{sub sample}}$** signal controls both gating sample-and-hold amplifiers.

Parameter	Description	Value
t_{L1S}	LED1 illuminated before LED1 sample initiated	33 µs
t_{L2S}	LED2 illuminated before LED2 sample initiated	33 µs
t_{OAS}	Both LEDs extinguished before ambient sample initiated	29 µs
t_{ASS}	Ambient sample initiated before subtraction samples	17 µs
t_{rest}	Time between LED1 extinguished and LED2 illuminated	8 µs
t_{1S}	LED1 sample duration	24 µs
t_{2S}	LED2 sample duration	24 µs
t_{SS}	Subtracted signals sample duration	18 µs
t_{AS}	Ambient sample duration	36 µs
t_{L1P}	LED1 illuminated pulse duration	58 µs
t_{L2P}	LED2 illuminated pulse duration	58 µs

Figure I.6 – Multiplexed timing parameters

Detailed numerical timing parameters for figure I.5 are shown in figure I.6. These values represent a 5.05kHz LED-switching frequency (at 30.3% duty cycle), achieved with a 4MHz (nominal) crystal oscillator for the PIC16C54. Variation in these parameters can only occur with oscillator frequency changes, resulting in a very high measured stability. In addition, any variations that do occur will have an identical effect on all signals, implying that timing signals will *track* each other; further stabilising the timing parameters.

Determination of the optimal timing parameters shown in figure I.6 required consideration of a number of characteristics of both the opto-electronics and the electronic circuitry. The major limiting factor in the speed of operation is the response time of the PiN diode amplifier to the step input produced by illuminating (extinguishing) an LED source. The time between an illumination change and the corresponding sample signal (t_{L1S}, t_{L2S}, t_{OAS}) must therefore allow time for the preamplifier to settle. Sample time for the separate signals (t_{1S}, t_{2S}, t_{AS}, t_{SS}) has a minimum defined by the sample-and-hold amplifiers, but in practice utilises the

maximum time available. Note also that the sample signals occur as early as possible to allow for settling of subsequent signal-processing stages before the initiation of analogue-to-digital conversion.

I.1.2b Analogue to digital conversion

In addition to analogue output of the ambient-subtracted signals, circuitry is included to provide sampled digital outputs. This sampling scheme also has four inputs that are used in conjunction with additional signal-processing stages to provide digital output of filtered and amplified versions of the signals. The six analogue inputs are multiplexed by an ADG408 analogue switch (IC6), which provides input to an ADS7805 analogue-to-digital converter (IC7). The ADS7805 is a 16-bit, 100KHz successive approximation register (SAR) ADC, including internal sample-and-hold circuitry, a stable reference voltage source and a complete digital output section, operating from a single 5V supply. It is therefore only necessary to sequentially select an input channel from the analogue switch, initiate a sample of that channel following a suitable settling time and, after the conversion is complete, output the digital word. By repeatedly cycling through all channels, a multiplexed digital output can provide information about all six channels.

The multiplexer (analogue switch) channel is determined by a 3-bit digital word, comprising the signals **A0**, **A1** and **A2**, which are generated by the micro-controller (IC11). Control of the ADC is by two digital signals, **R/$\overline{\mathbf{C}}$** and $\overline{\mathbf{CS}}$. The signal **R/$\overline{\mathbf{C}}$** determines whether the ADC should **R**ead (output the previously sampled value) or **C**onvert (initiate the next conversion), whilst $\overline{\mathbf{CS}}$ (**C**hip **S**elect) is negatively asserted or pulsed to begin the selected operation. One additional signal is also provided by the micro-controller, a **D**ata **V**alid (**DV**) output that indicates, via a rising edge, that output data from the ADC is valid. The exact timing of the digital sampling scheme is shown in figure I.7, with signals generated by the micro-controller highlighted. The $\overline{\textbf{sub sample}}$ gating control signal is included to demonstrate the synchronisation between the sampling scheme timing and the multiplexing scheme timing shown previously (figure I.5).

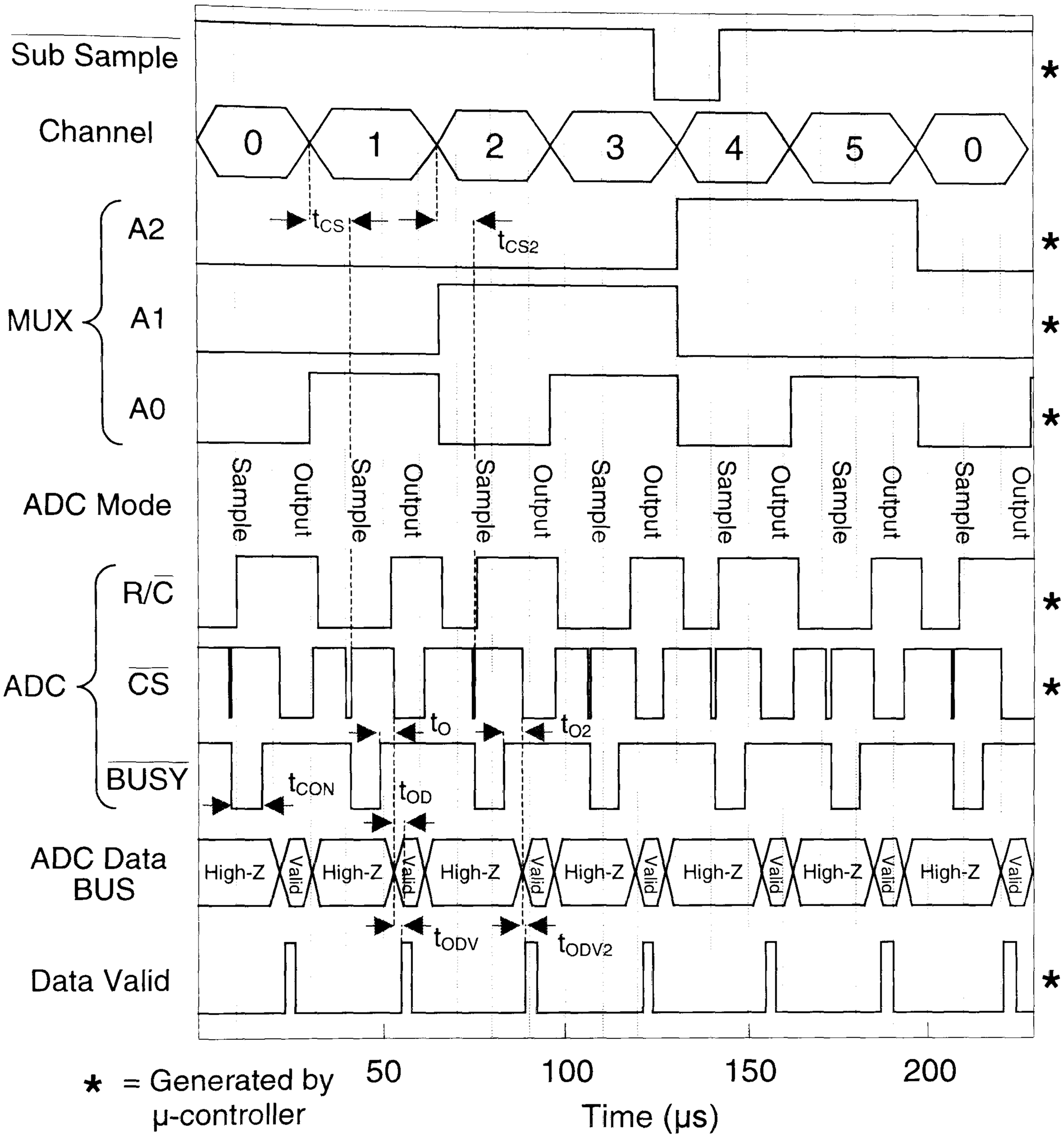


Figure I.7 – Digital sampling scheme timing

Also shown in the timing diagram of figure I.7 is the state of the ADC output data-bus and the ADC output signal **BUSY**. The **BUSY** output is negatively asserted by the ADC to indicate that a conversion is in progress. Although this signal is not directly used in this implementation, it is instructive to include it in the timing diagram, where the pulse lengths shown demonstrate the maximum possible conversion time. Detailed numerical timing values for the scheme of figure I.7 are shown in figure I.8.

Parameter	Description	Value
t_{CS}/t_{CS2}	Multiplexer channel selected before sample initiated	11/10 μs
t_{CON}	ADC Conversion time	8 μs (max)
t_O/t_{O2}	Sample complete before ADC read command	4/5 μs (min)
t_{OD}	ADC output valid after read command	80 ns (max)
t_{ODV}/t_{ODV2}	Data valid rising edge after read command	2/1 μs

Figure I.8 – Sampling scheme timing parameters

I.1.2c Digital output

Output from the ADC is buffered by a pair of non-inverting bus drivers (IC8 & 9) that provide the overall digital output of the circuit. The digital interface is completed by the addition of timing signals that indicate the current channel, **A0**, **A1** and **A2**, and a pulse whose rising edge can be used to clock valid data into an external circuit, **DV**. In addition, a 3-to-8 address decoder (IC12) is used to fully decode the current-channel signals, and by enabling the chip output with the data valid signal, channel-dependent data valid signals are produced. This means that an external latched circuit can directly connect to the digital output and can isolate any channel by using one of these signals as a data clock.

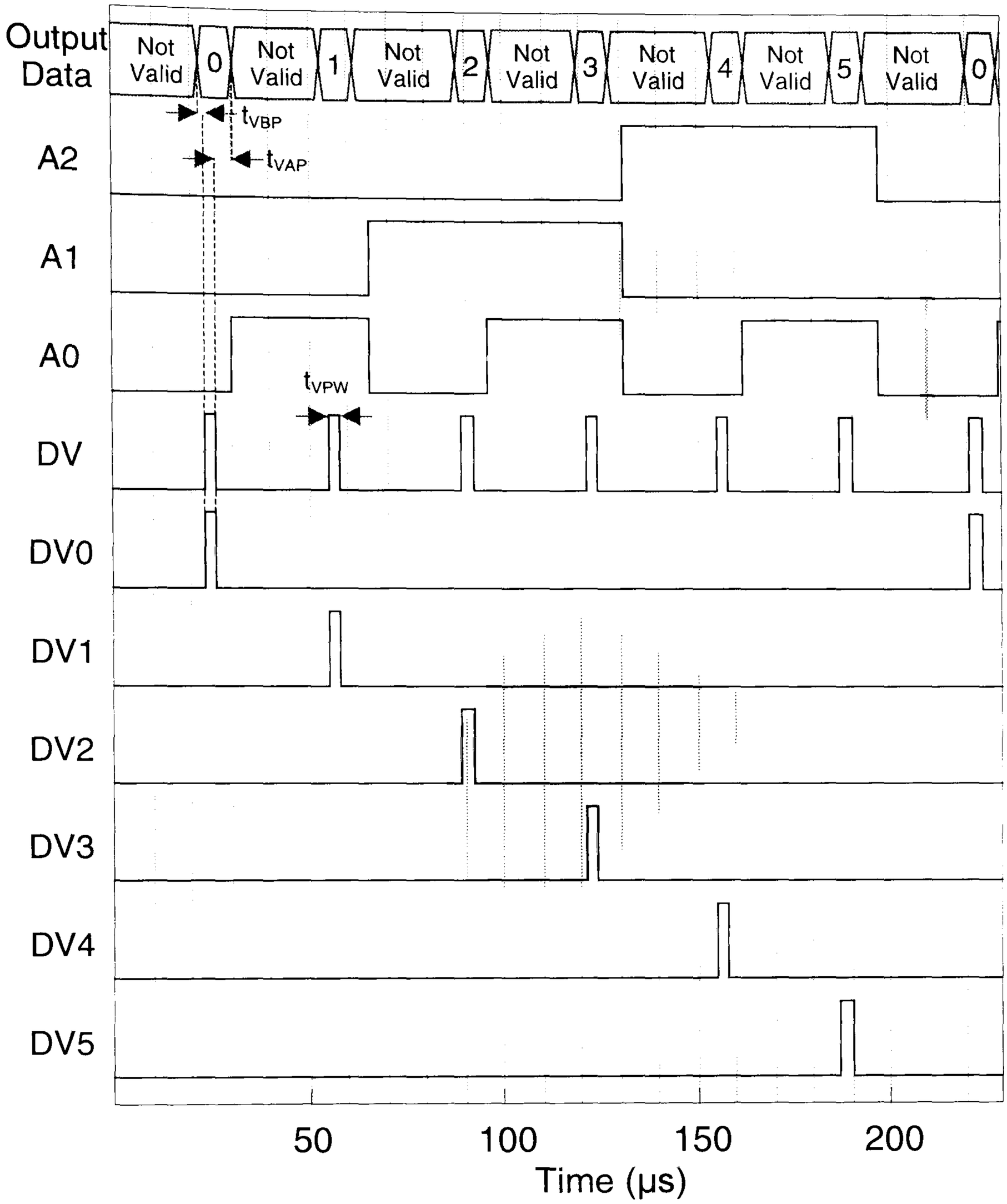


Figure I.9 – Digital output scheme timing

Figure I.9 shows the timing for the digital output scheme, synchronised to the same time reference as figures I.5 and I.7. The numerical values of timing parameters needed to sucessfully implement an interface with an external circuit using this output scheme are shown in figure I.10.

Parameter	Description	Value
t_{VBP}	Output data valid before rising edge of data valid pulse	1 µs (min)
t_{VAP}	Output data valid after falling edge of data valid pulse	3 µs (min) 5 µs (max)
t_{VPW}	Data valid pulse width (any data vaid signal)	3 µs (nom)

Figure I.10 – Output scheme timing parameters

I.1.2d Software Implementation

The actual micro-controller code used to generate the circuit timing is reproduced for reference in figures I.11 and I.12. The mapping between output pins of the micro-controller and the previously assigned functions is detailed in the software listing, as is the use of an internal watchdog timer to ensure proper power-up operation and that the circuitry can recover adequately from transient interruptions in the power supply. The detailed timing is produced by the fixed length of an instruction cycle, and this implementation will therefore only produce the correct timing when used with a 4MHz crystal.

```
;       *********************
;       * Signal Generator *
;       *  for PPG Probe   *
;       *  ---PIC1654---   *
;       * Author: M. Hayes *
;       * LU   Mar 98  OEG *
;       *********************

;Version 6.7      - Logic Sequence Generator
;PIC16C54 - 4MHz XTAL OSC
;Required output for LEDs, S&H and MUX on PORTB
;with SUB SAMPLE, DV, R/C and CS on PORTA
;(Data Valid, Read / Convert and Chip Select)
;Approx 5.05kHz per LED Operation
;Sampler works at 30.3kHz over 6 channels
;       ---PORTB---
;       Bit7 = LED1
;       Bit6 = LED2
;       Bit5 = Led1 Sample (NOT)
;       Bit4 = Led2 Sample (NOT)
;       Bit3 = Ambient Sample (NOT)
;       Bit2 = A0 / S0 (MUX)
;       Bit1 = A1 / S1 (MUX)
;       Bit0 = A2 / S2 (MUX)
;       ----------
;       ---PORTA---
;       Bit3 = Sub Sample (NOT)
;       Bit2 = DV (Rising Edge = Data Valid)
;       Bit1 = R/C (NOT)
;       Bit0 = CS (NOT)
;       ----------
;       Watchdog timer support included

;*---General Equates---*
RTCC    EQU 1    ;Counter
PC      EQU 2    ;Program Counter
STATUS  EQU 3    ;Status Register
W       EQU 0    ;Result Destination is W Register
F       EQU 1    ;Result Destination is F Register
PORTA   EQU 5    ;4 Bit Bidirectional Port A
PORTB   EQU 6    ;8 Bit Bidirectional Port B

;*---Variable and Pinout Definitions---*
#define  LED1     PORTB,7
#define LED2      PORTB,6
#define  L1S      PORTB,5
#define  L2S      PORTB,4
#define  AS       PORTB,3
#define  A0       PORTB,2
#define  A1       PORTB,1
#define  A2       PORTB,0
#define  SUBS     PORTA,3
#define  DV       PORTA,2
#define  RC       PORTA,1
#define  CS       PORTA,0
CLEAR   EQU      0F8H
ZERO    EQU      0H
ONE     EQU      04H
TWO     EQU      02H
THREE   EQU      06H
FOUR    EQU      01H
FIVE    EQU      05H

;*Reset*
ORG 000H
START   MOVLW 09H        ;W=09
        OPTION           ;Prescaler = WDT / 2
;*-------------WDT Notes-------------------*
;*Increase Prescaler value until WDT Timeout*
;*Does not occur in simulation.  WDT will   *
;*then cause RESET if program gets stuck in *
;*  loop, eg. with a power spike/surge      *
;-----------------------------------------
        CLRWDT           ;Reset WDT
        MOVLW   0H       ;Store 0 in W
        TRIS    PORTB    ;Make PortB all Outputs
        TRIS    PORTA    ;Make PortA all Outputs
        MOVLW   038H     ;All signals deasserted
        MOVWF   PORTB    ;Reset Outputs
        MOVLW   0BH      ;All signals deasserted
        MOVWF   PORTA    ;Reset Outputs
        BSF     LED1     ;Turn ON LED1
        GOTO    ENTRY    ;Start at entry point

;*---Main Program Loop---*
MAIN

;************************
;*---LED1 Time Period---*
;*    66uS Execution    *
;************************
        NOP              ;Delay
        NOP              ;To Allow PiN Amp
        NOP              ;to settle properly
        BCF     CS       ;Initiate Conversion
        BSF     CS       ;(1uS Pulse)
ENTRY   NOP
        NOP
        NOP
        NOP              ;Wait for 7uS
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        BSF     RC       ;ADC Output
        BCF     CS       ;Chip Select Initiates
        BCF     L1S      ;Assert Sample
        BSF     DV       ;Data Valid
        MOVF    PORTB,W  ;Get PORTB Value
        ANDLW   CLEAR    ;Clear Address Bits
        BCF     DV       ;Data Valid End
        NOP              ;No Need to Sub-Sample
        IORLW   ZERO     ;Channel 0
        BSF     CS       ;Chip Select Ends Out
        NOP
        NOP
        NOP
        MOVWF   PORTB    ;Select MUX Channel
        BCF     RC       ;ADC Convert
        CLRWDT           ;Clear Watchdog timer
        NOP              ;Wait 3uS to allow
        NOP              ;MUX to settle
        NOP
        NOP              ;Wait 6uS for ADC to
        NOP              ;Complete Conversion
        NOP
        BCF     CS       ;Initiate Conversion
        BSF     CS       ;(1uS Pulse)
        BSF     RC       ;Next Initiation Out
        BSF     SUBS     ;De-assert Sub-Sample
        BSF     L1S      ;De-assert Sample
        BCF     LED1     ;Turn OFF LED
        NOP              ;Wait for 7uS
        NOP
        NOP
        NOP
        NOP
        NOP
        BSF     LED2     ;Turn ON LED
        BCF     CS       ;Initiate Output
        BSF     DV       ;Data Vaid
        MOVF    PORTB,W  ;Get PORTB Value
        ANDLW   CLEAR    ;Clear Address Bits
        BCF     DV       ;Data Valid End
        IORLW   ONE      ;Channel 1
        NOP              ;Delay 2us
        NOP
        MOVWF   PORTB    ;Select MUX Channel
```

Figure I.11 – Micro-controller code (Part 1)

```
        BSF     CS      ;Chip Select Ends Output
        BCF     RC      ;ADC Convert
        NOP             ;No Loop
        NOP             ;So Delay
        NOP
        NOP

;************************
;*---LED2 Time Period---*
;*    66uS Execution    *
;************************
        NOP             ;Delay
        NOP             ;To Allow PiN Amp
        NOP             ;to settle properly
        BCF     CS      ;Initiate Conversion
        BSF     CS      ;(1uS Pulse)
        NOP
        NOP
        NOP
        NOP             ;Wait for 7uS
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        BSF     RC      ;ADC Output
        BCF     CS      ;Chip Select Initiates
        BCF     L2S     ;Assert Sample
        BSF     DV      ;Data Valid
        MOVF    PORTB,W ;Get PORTB Value
        ANDLW   CLEAR   ;Clear Address Bits
        BCF     DV      ;Data Valid End
        NOP             ;No Need to Sub-Sample
        IORLW   TWO     ;Channel 2
        BSF     CS      ;Chip Select Ends Output
        NOP
        NOP
        NOP
        MOVWF   PORTB   ;Select MUX Channel
        BCF     RC      ;ADC Convert
        CLRWDT          ;Clear Watchdog timer
        NOP             ;Wait 3uS to allow
        NOP             ;MUX to settle
        NOP
        NOP             ;Wait 6uS for ADC to
        NOP             ;Complete Conversion
        NOP
        BCF     CS      ;Initiate Conversion
        BSF     CS      ;(1uS Pulse)
        BSF     RC      ;Next Init Outputs
        BSF     SUBS    ;De-assert Sub-Sample
        BSF     L2S     ;De-assert Sample
        BCF     LED2    ;Turn OFF LED
        NOP             ;Wait for 7uS
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP             ;No LEDs ON
        BCF     CS      ;Initiate Output
        BSF     DV      ;Data Vaid
        MOVF    PORTB,W ;Get PORTB Value
        ANDLW   CLEAR   ;Clear Address Bits
        BCF     DV      ;Data Valid End
        IORLW   THREE   ;Channel 3
        NOP             ;Delay 2us
        NOP
        MOVWF   PORTB   ;Select MUX Channel
        BSF     CS      ;Chip Select Ends Out
        BCF     RC      ;ADC Convert
        NOP             ;No Loop
        NOP             ;So Delay
        NOP
        NOP
        NOP

;*****************************
;*---Ambient Time Period---*
;*      66uS Execution     *
;*****************************
        NOP             ;Delay
        NOP             ;To Allow PiN Amp
        NOP             ;to settle properly
        BCF     CS      ;Initiate Conversion
        BSF     CS      ;(1uS Pulse)
        BCF     AS      ;Assert Sample
        NOP             ;Wait for 7uS
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        BSF     RC      ;ADC Output
        BCF     CS      ;Chip Select Initiates
        NOP
        BSF     DV      ;Data Valid
        MOVF    PORTB,W ;Get PORTB Value
        ANDLW   CLEAR   ;Clear Address Bits
        BCF     DV      ;Data Valid End
        BCF     SUBS    ;Assert Sub-Sample
        IORLW   FOUR    ;Channel 4
        BSF     CS      ;Chip Select Ends Out
        NOP
        NOP
        NOP
        MOVWF   PORTB   ;Select MUX Channel
        BCF     RC      ;ADC Convert
        CLRWDT          ;Clear Watchdog timer
        NOP             ;Wait 3uS to allow
        NOP             ;MUX to settle
        NOP
        NOP             ;Wait 6uS for ADC to
        NOP             ;Complete Conversion
        NOP
        BCF     CS      ;Initiate Conversion
        BSF     CS      ;(1uS Pulse)
        BSF     RC      ;Next Initiation Out
        BSF     SUBS    ;De-assert Sub-Sample
        BSF     AS      ;De-assert Sample
        NOP             ;No LEDs to turn off
        NOP             ;Wait for 7uS
        NOP
        NOP
        NOP
        NOP
        NOP
        NOP
        BSF     LED1    ;Turn ON LED
        BCF     CS      ;Initiate Output
        BSF     DV      ;Data Vaid
        MOVF    PORTB,W ;Get PORTB Value
        ANDLW   CLEAR   ;Clear Address Bits
        BCF     DV      ;Data Valid End
        IORLW   FIVE    ;Channel 5
        NOP             ;Delay 2us
        NOP
        MOVWF   PORTB   ;Select MUX Channel
        BSF     CS      ;Chip Select Ends Out
        BCF     RC      ;ADC Convert
        NOP
        NOP
        GOTO    MAIN    ;Loop round again

;*------------Reset-----------*
        ORG 1FFH
        GOTO START

        END
```

Figure I.12 – Micro-controller code (Part 2)

I.1.3 Implementation details

This section describes various other issues in the detailed implementation of the front-end circuitry, highlighting several important aspects of the overall design methodology.

I.1.3a Power supply

Overall power to the circuit is provided by an externally regulated ±12V supply, which directly powers most analogue components. A local 5V regulator (IC22 and associated circuitry) provides power for the low-voltage logic chips and micro-controller. An additional 5V regulator (IC10 and associated circuitry) is also utilised to directly supply the analogue-to-digital converter. This methodology assures stability of the ADC internal voltage reference, which could otherwise be affected by power-supply noise induced by the high-speed switching of the logic circuitry.

I.1.3b External connectors

The main power-supply connector (PL6) accepts ±12V and analogue-ground from an external source and provides a digital-ground for connection to external logic circuitry, although these functions are also reproduced on other connectors for convenience (PL1). The digital output connector (PL4 along with a few spare pins of PL6) encompasses all buffered digital output from the sampling scheme, as well as all output timing signals (including the decoded data-valid outputs). The analogue connector (PL1) provides analogue output of the ambient-subtracted pair of signals and four analogue inputs to the sampling scheme. The connector is arranged so that, in conjunction with filtering and amplification circuitry described later, jumpers between pins of this connector can select either the built-in signal-processing stages or other undefined external stages. This enables much greater flexibility in both testing and operation of the circuit.

I.2 FILTERING AND AMPLIFICATION

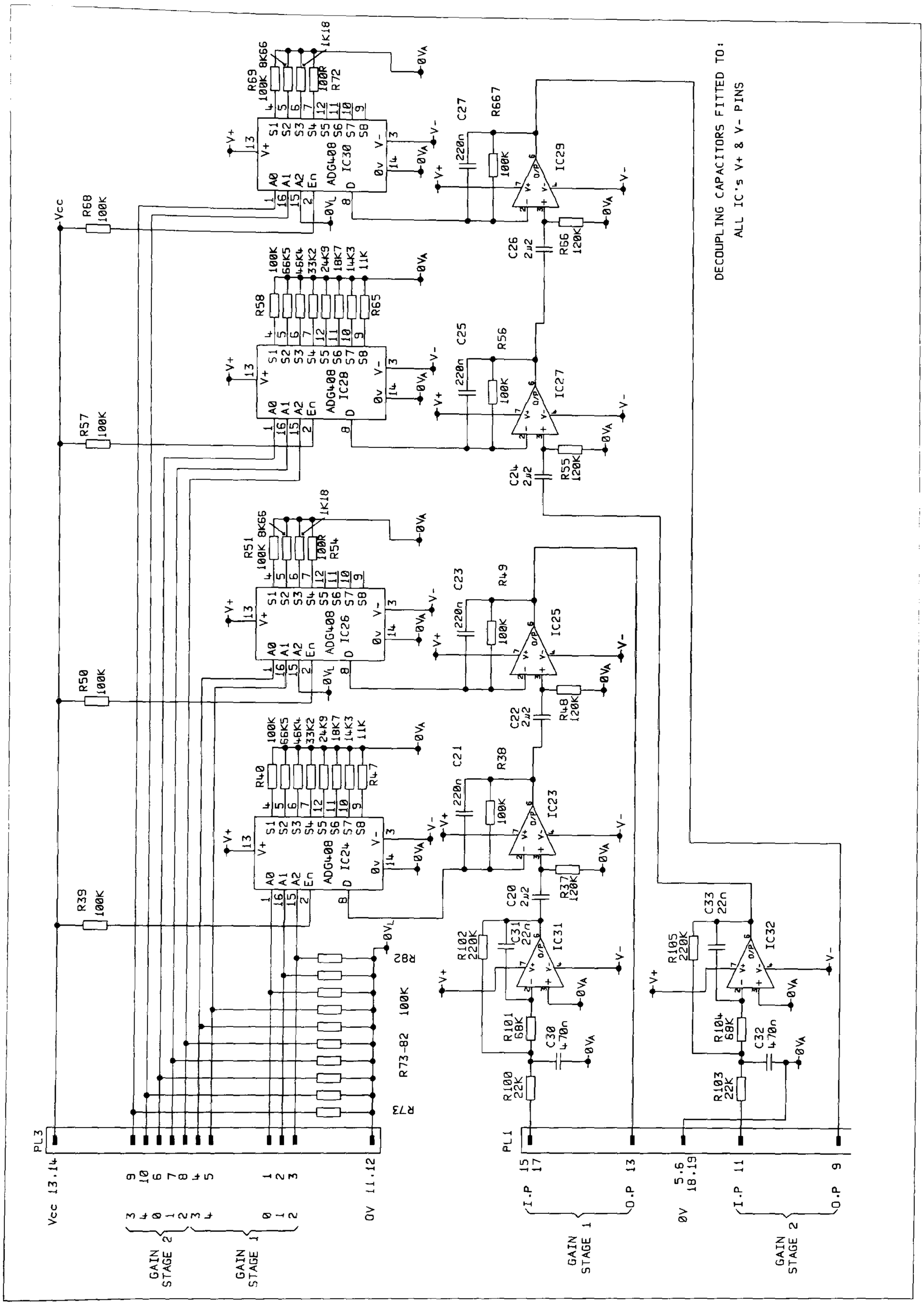


Figure I.13 – Schematic of filtering and amplification circuitry

Filtering and amplification stages are included to provide the necessary signal processing to isolate the arterial PPG signal and maintain adequate size signals for the sampling scheme. These stages can be enabled or disabled by jumper control, and the level of amplification applied to the signals can be controlled externally by means of a digital interface.

I.2.1 Pre-processing filters

Initially, signals entering the filtering and amplification stages are pre-processed by low-pass filters. These filters apply limited gain to the signals, whilst attenuating any high-frequency components such as switching noise or high-frequency interference. The cut-off frequency for these filters lies outside the expected spectra of PPG signals, and therefore does not significantly affect the desired information.

I.2.1a Multiple feedback filter

The low-pass filters used are based upon a multiple-feedback topology, as shown in figure I.14. The voltage transfer function of this circuit in the *laplace* domain is given by

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{R_4}{s^2 C_5 R_1 R_3 R_4 + sC_5\left(R_1 R_3 + R_1 R_4 + R_3 R_4\right) + R_1}, \qquad \textbf{[I.6]}$$

where the frequency response is determined by making the substitution $s = j\omega$.

I.2.1b Implementation

Examination of equation [I.6] demonstrates that this is a low-pass filter with a DC gain of R_4/R_1. The component values used in this circuit were $R_1 = 22k\Omega$, $R_3 = 68k\Omega$, $R_4 = 220k\Omega$, $C_2 = 470nF$ and $C_5 = 22nF$. This produces a linear-phase (Bessel) frequency response with a cut-off frequency of 10.2Hz and a DC gain of 10 (20dB), as demonstrated in the bode plots of figure I.15. The phase response also serves to emphasise the output inversion inherent in this type of filter circuit.

The pre-processing filters are implemented using OP177 op-amps (IC31 & 32), which exhibit high precision and low drift.

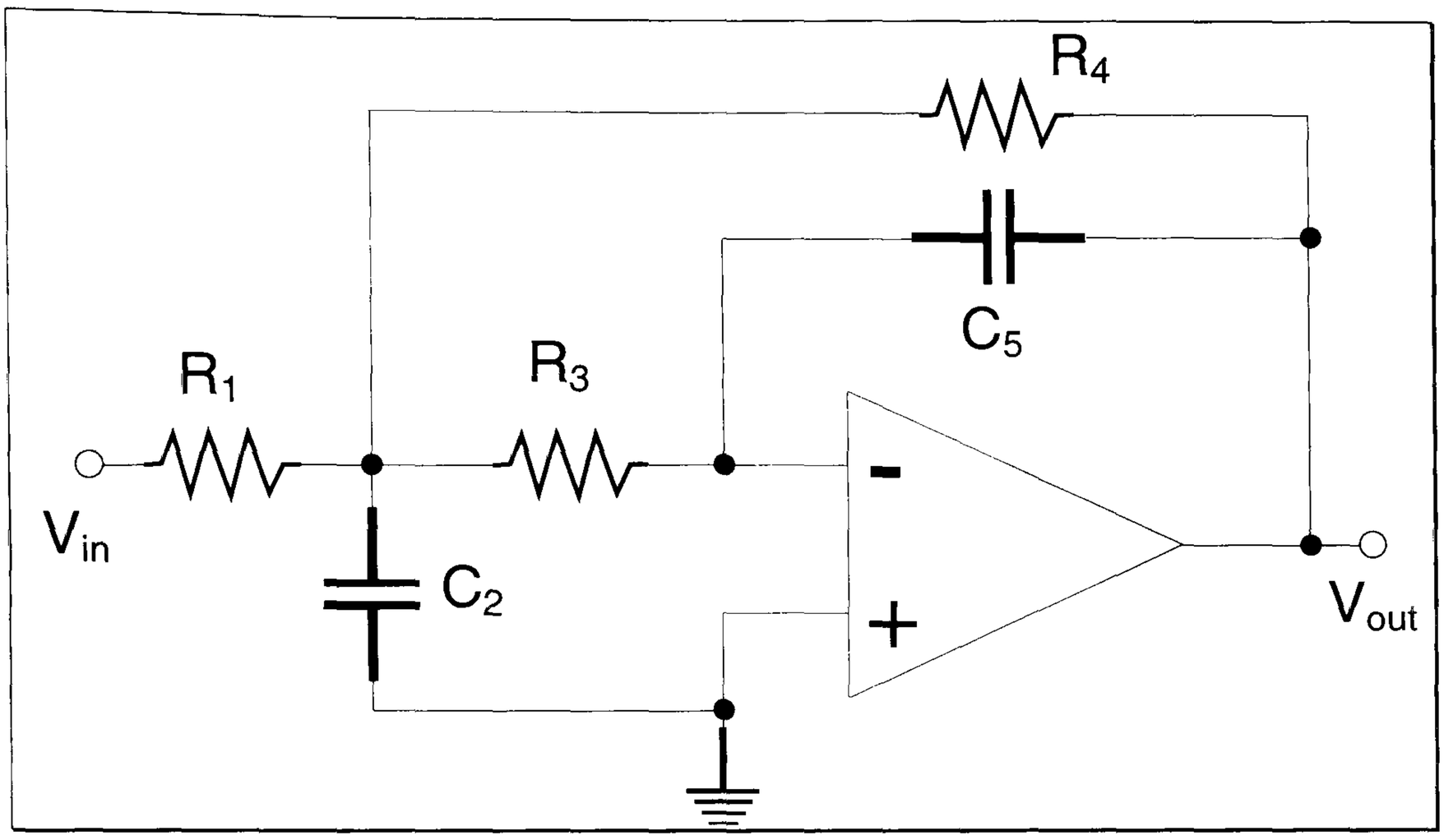


Figure I.14 - Multiple feedback low-pass filter topology

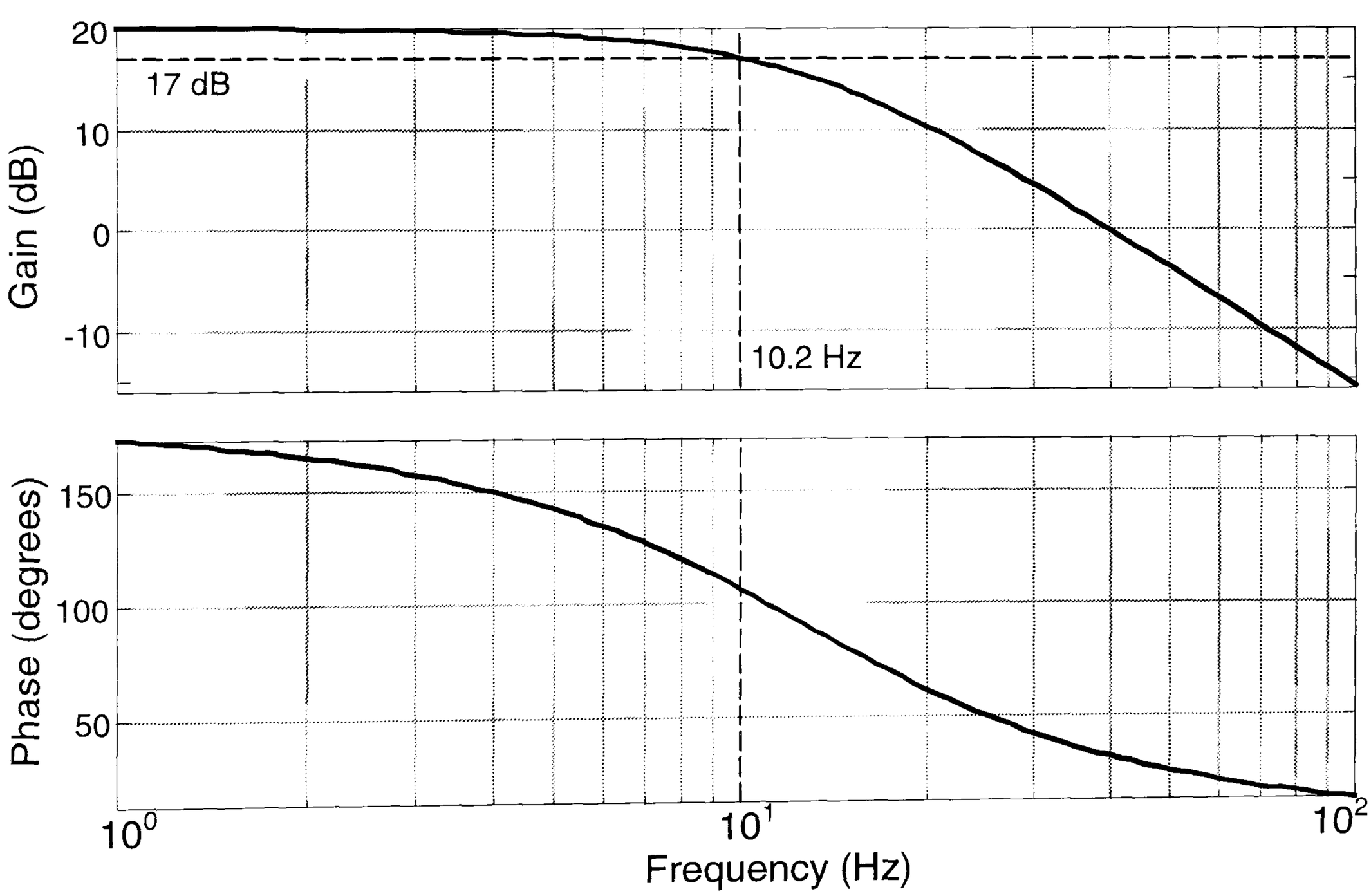


Figure I.15 – Frequency response of pre-processing filters

I.2.2 Amplifying filters

Following the pre-processing filters, circuitry is provided to selectively amplify the desired frequency components. This circuitry takes the form of a cascaded pair of amplifiers whose gains may be controlled by an external digital interface and whose frequency response is chosen to amplify a spectral region extending from about 0.75Hz to 5Hz, encompassing the expected spectra of PPG signals. By providing external control of the applied gain, it is possible to observe an extremely wide dynamic range of possible PPG signals. This dynamic range is further enhanced by producing a logarithmic gain characteristic, with a resolution of 2dB per digital bit.

I.2.2a Common amplification stage

The complete filtering and amplification stages are constructed by cascading a pair of basic band-limited amplification circuits. These basic circuits filter out any DC component, whilst selectively amplifying the desired spectral range. The schematic for such a stage is shown in figure I.16, resulting in a voltage transfer function given by

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{s^2 R_1 R_3 R_4 C_2 C_5 + sR_1 C_2 (R_3 + R_4)}{s^2 R_1 R_3 R_4 C_2 C_5 + sR_3 (R_1 C_2 + C_5 R_4) + R_3}. \qquad \textbf{[I.7]}$$

Examination of equation [I.7] reveals that this circuit will block DC signals, acting as a high-pass filter with a cut-off frequency of $1/2\pi R_1 C_2$ Hz. At frequencies slightly above this lower cut-off frequency, the circuit will display a gain of A dB, where

$$\mathrm{A} \approx 20 \log_{10}\left(1 + \frac{R_4}{R_3}\right). \qquad \textbf{[I.8]}$$

However, this gain will start to decrease for frequencies above a higher cut-off frequency of $1/2\pi R_4 C_5$ Hz, until, for very high frequencies, the gain is reduced to unity (0dB). Since neither the lower or upper cut-off frequency is dependent upon

R_3, modification of this value modifies the gain only and can be used to dynamically control the degree of amplification.

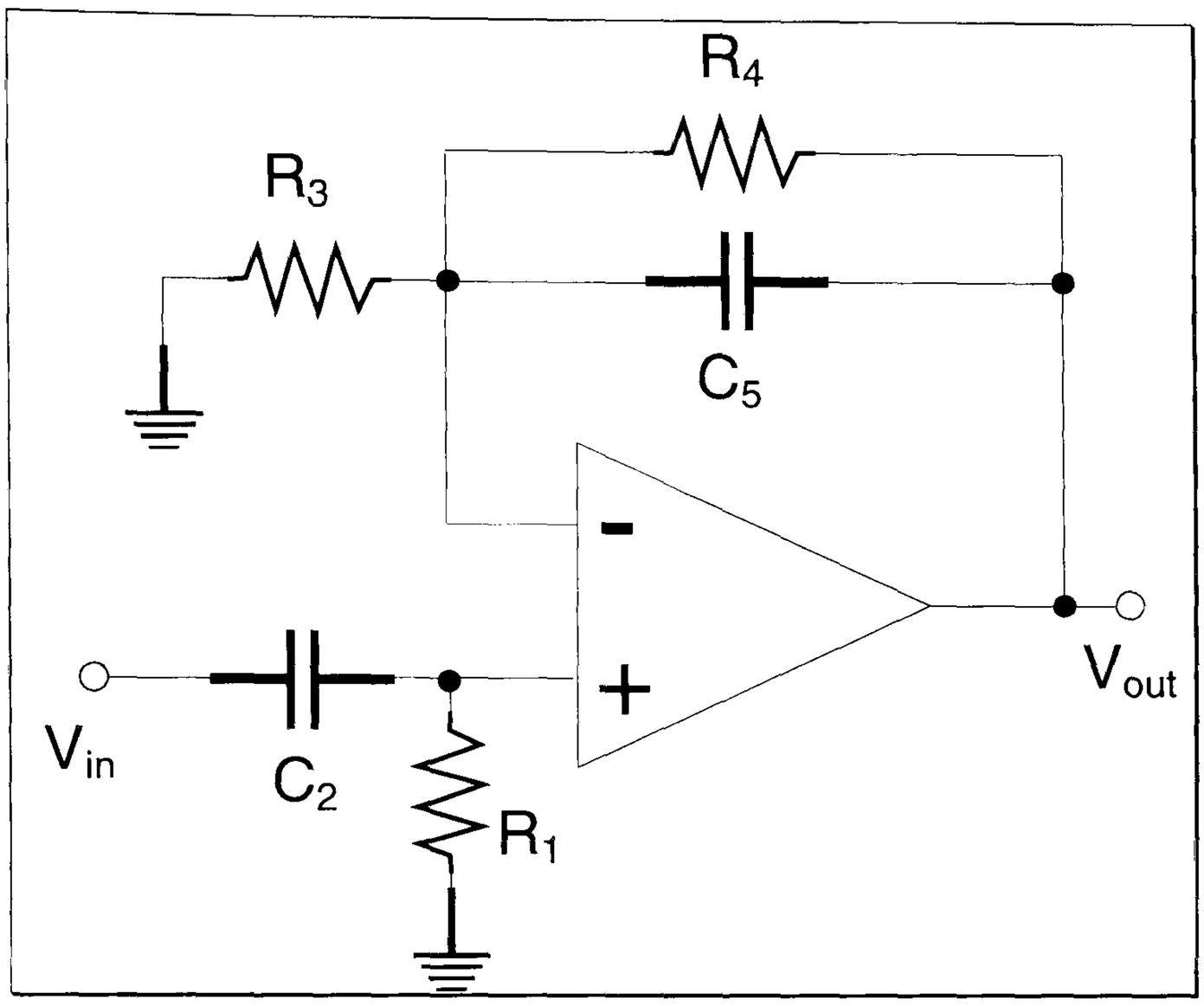


Figure I.16 - Common amplification stage

The use of this circuit configuration not only enables straightforward modification of the gain using only a single variable resistor, but also displays an extremely linear phase response in the pass-band (see figure I.17). Although the gain level in the pass-band is not maximally flat, experience dictates that accurate recovery of PPG signals and comparison between them is more sensitive to phase errors than gain variations. In addition, the use of such limited-feedback techniques greatly simplifies cascading of such circuits, without the complex phase relations that can occur from combinations of other filters (e.g. Bessel, Butterworth, etc.). Although this methodology results in a relatively soft transition from pass-band to stop-band, this does not impose a severe performance reduction in the current application.

I.2.2b Implementation of common stage

The component values used to implement this circuit were $R_1 = 120k\Omega$, $R_4 = 100k\Omega$, $C_2 = 2.2\mu F$ and $C_5 = 220nF$, with R_3 used to determine the circuit gain. These values produce a lower cut-off frequency of around 0.6Hz and an upper cut-off

frequency of around 7.2Hz, with practical gains realisable from around 6dB ($R_3 = 100k\Omega$) to around 60dB ($R_3 = 100\Omega$). The frequency response for this circuit is shown in figure I.17 with a 10dB gain selected ($R_3 \approx 46k\Omega$).

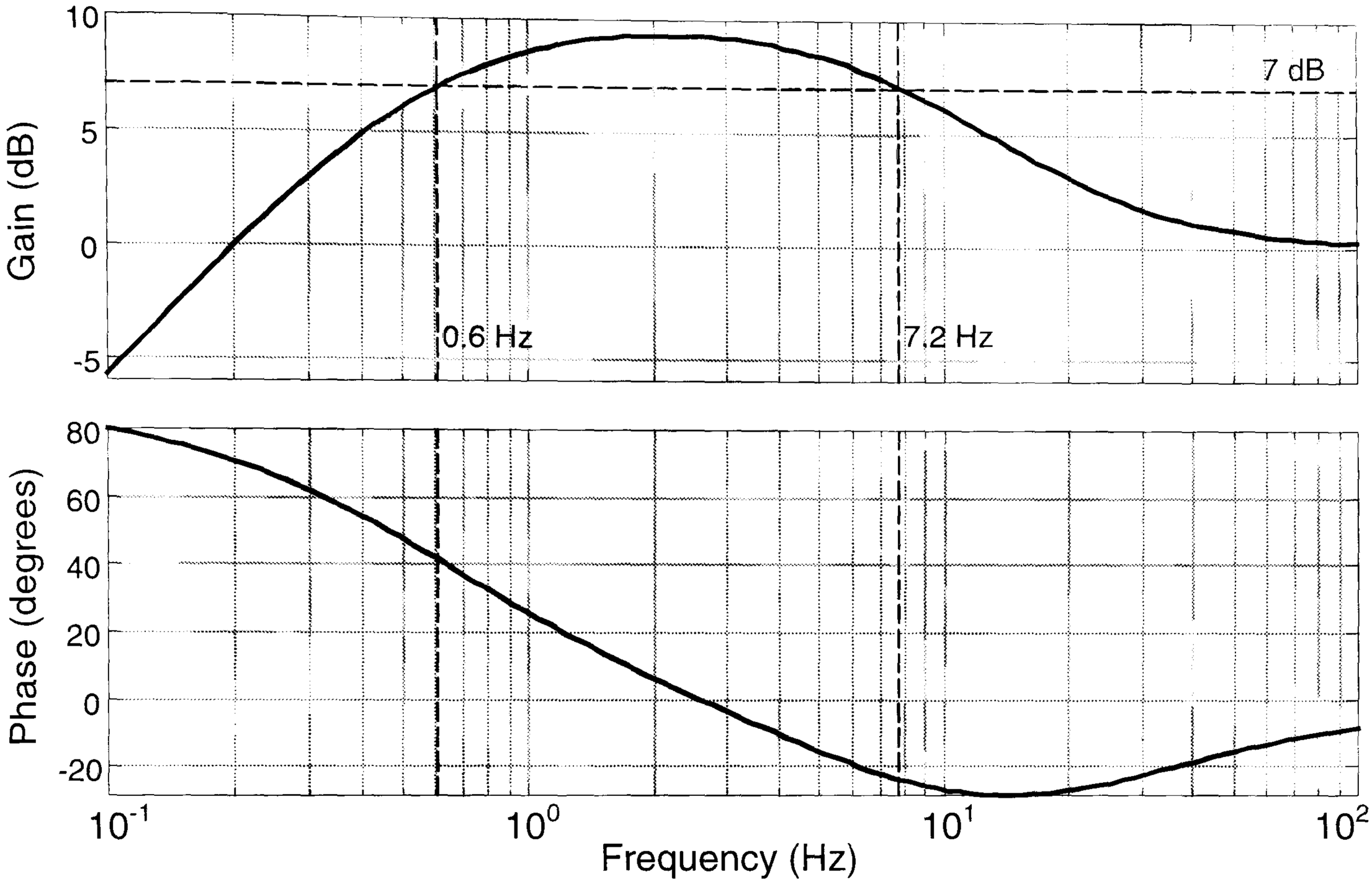


Figure I.17 – Frequency response of common amplification stage

OP177 op-amps were again used in the practical implementation of these circuit blocks (IC23, 25, 27 and 29). Selection of a particular gain is accomplished by switching different resistors into the circuit, therefore modifying the value of R_3. ADG408 analogue multiplexers (IC24, 26, 28 and 30) perform the switching under control of an external digital interface. These devices display an on-resistance of around 150Ω, which must be considered when calculating the necessary resistance for a particular gain.

I.2.2c Cascaded amplifying filter

Cascading a pair of common amplifying stages, with the gain characteristics of the two stages chosen to produce a combined continuous logarithmic characteristic,

results in a complete amplifying filter. Whilst the first stage employs 8 different gains, the second stage uses only 4, giving 32 possible combined gain levels and requiring 5-bit digital control. The resistor values shown in figure I.13 (e.g. R40 to 47 and R51 to 54) produce gains ranging from 12dB to 72dB with a monotonic 2dB gain increase for every 1-bit increase in the digital control word. Because the analogue multiplexers have digital inputs to select their state, it is then only necessary to connect these inputs to a digital connector to complete the digitally controlled amplifying and filtering stages. The frequency response of a cascaded amplifying filter is shown in figure I.18, with an overall gain of 22dB (resistor values 18.7$k\Omega$ and 100$k\Omega$, digital word = 00110), demonstrating the increased stop-band attenuation and persistence of the linear phase. Note also that the combination of stages has resulted in the –3dB points being shifted to approximately 0.9Hz and 4.8Hz, since the original cut-off points (0.6Hz and 7.2Hz) are now 6dB less than the combined gain.

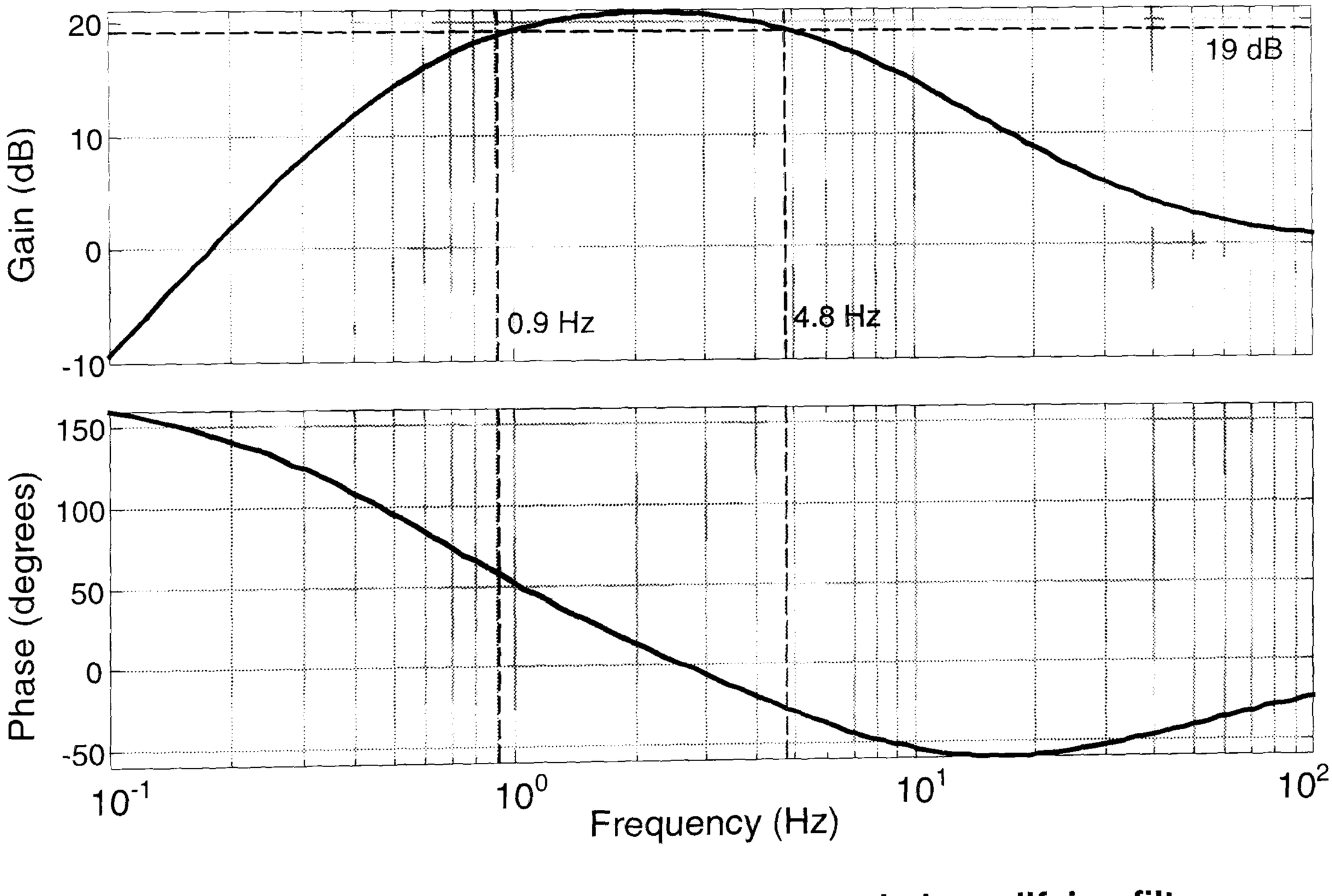


Figure I.18 – Frequency response of cascaded amplifying filter

I.2.3 Complete signal processing

Connecting the pre-processing stages to the amplifying filters completes the signal processing chain, providing controllable amplification of desired signal components and attenuation of undesired components. The frequency response for the overall circuit is shown in figure I.19, normalised with respect to the total circuit gain. Note that although the normalised frequency response varies with the selected gain, this has little effect in the pass-band. In addition, a highly linear phase response is observed for all gains. The degree of amplification possible with this circuit configuration varies from around 32dB to 100dB, enabling an extremely wide range of PPG sizes to be amplified to suitable levels for the sampling scheme.

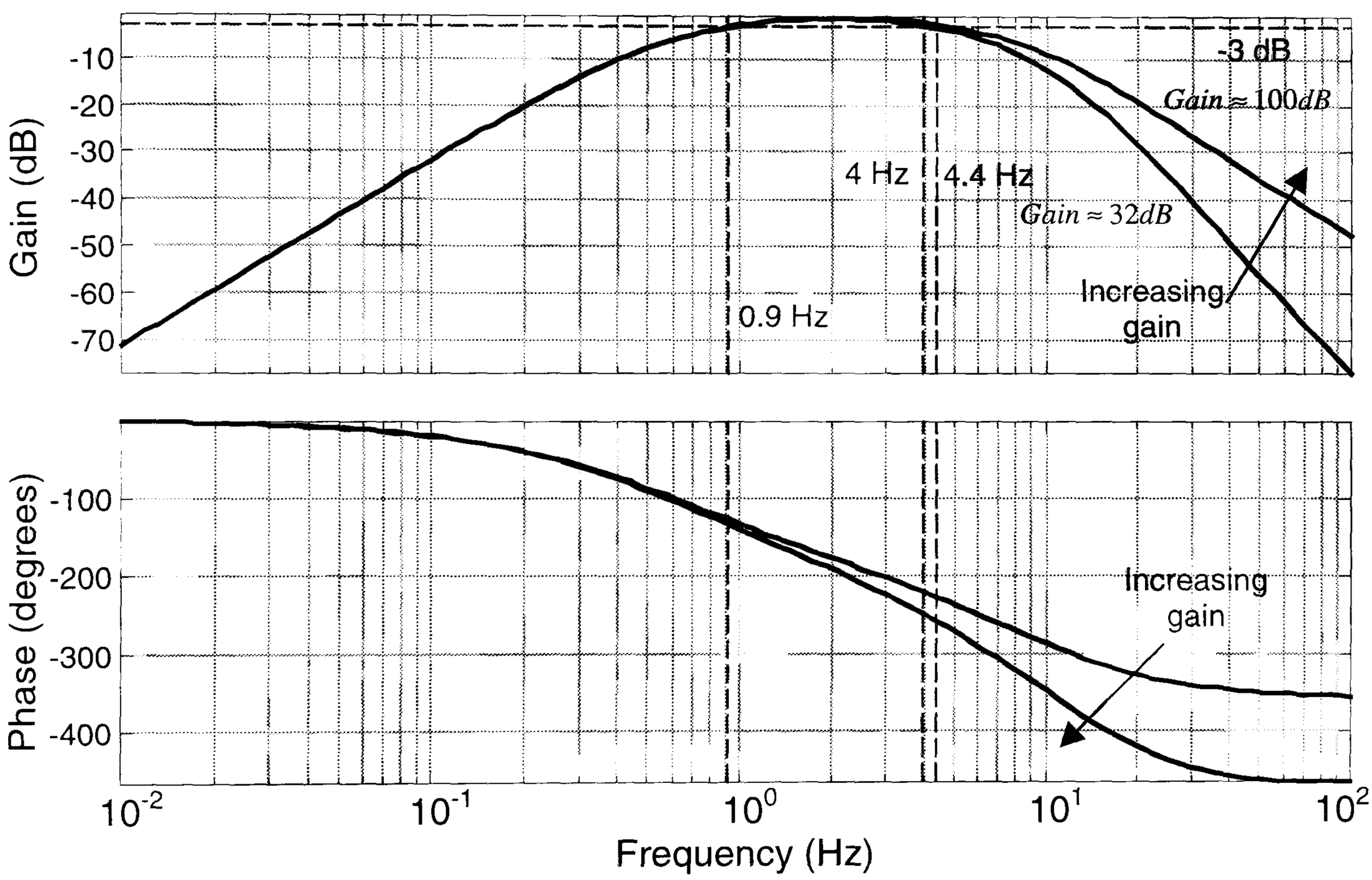


Figure I.19 – Normalised frequency response of complete circuit

I.3 LED DRIVERS

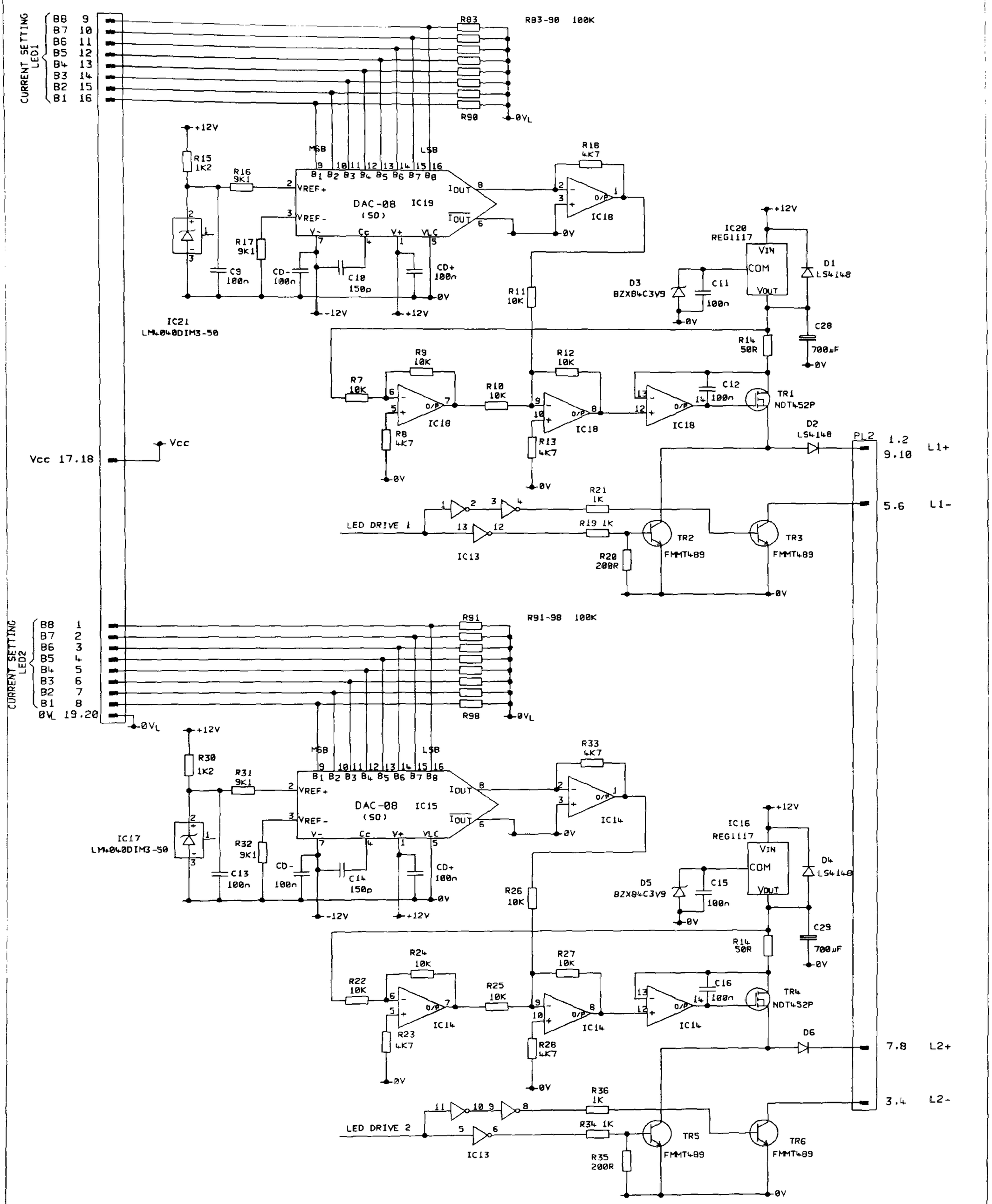


Figure I.20 – Schematic of LED driver circuitry

A pair of digitally controlled LED drivers implements the source illumination when instructed by the front-end circuitry. Precise control of the electrical power ensures that each LED provides the same level of brightness each time it is illuminated, whilst compound switching techniques reduce the switching noise induced by the circuitry. The LED drivers are capable of working independently to drive separate devices, or in combination to drive back-to-back diodes.

I.3.1 Methodology

Since the LEDs used in most commercial probes do not provide a means to independently measure the optical power, the only practical way to ensure a constant brightness is to control the electrical power supplied to the devices. The optical power output of an LED is proportional to the current flowing through it, and therefore this requirement equates to control of the LED current. In a simplistic situation this can be achieved by biasing the LED with a resistor, in a similar manner to the PiN diode biasing described in section I.1.1b. Unfortunately this method will only provide a constant current if both the LED voltage drop and the supply voltage remain constant, ultimately limiting the constancy of current and therefore the optical power. A better and more stable solution is an active circuit that delivers a constant and predefined current regardless of the LED voltage.

I.3.1a Current control

An active current control methodology is shown in figure I.21, which utilises direct measurement of the delivered current in a feedback scheme that renders the load current independent of the voltage across it. This not only ensures constancy of the LED power under switching conditions, but also enables the circuit to drive a wide range of physical devices without modification and with identical current levels.

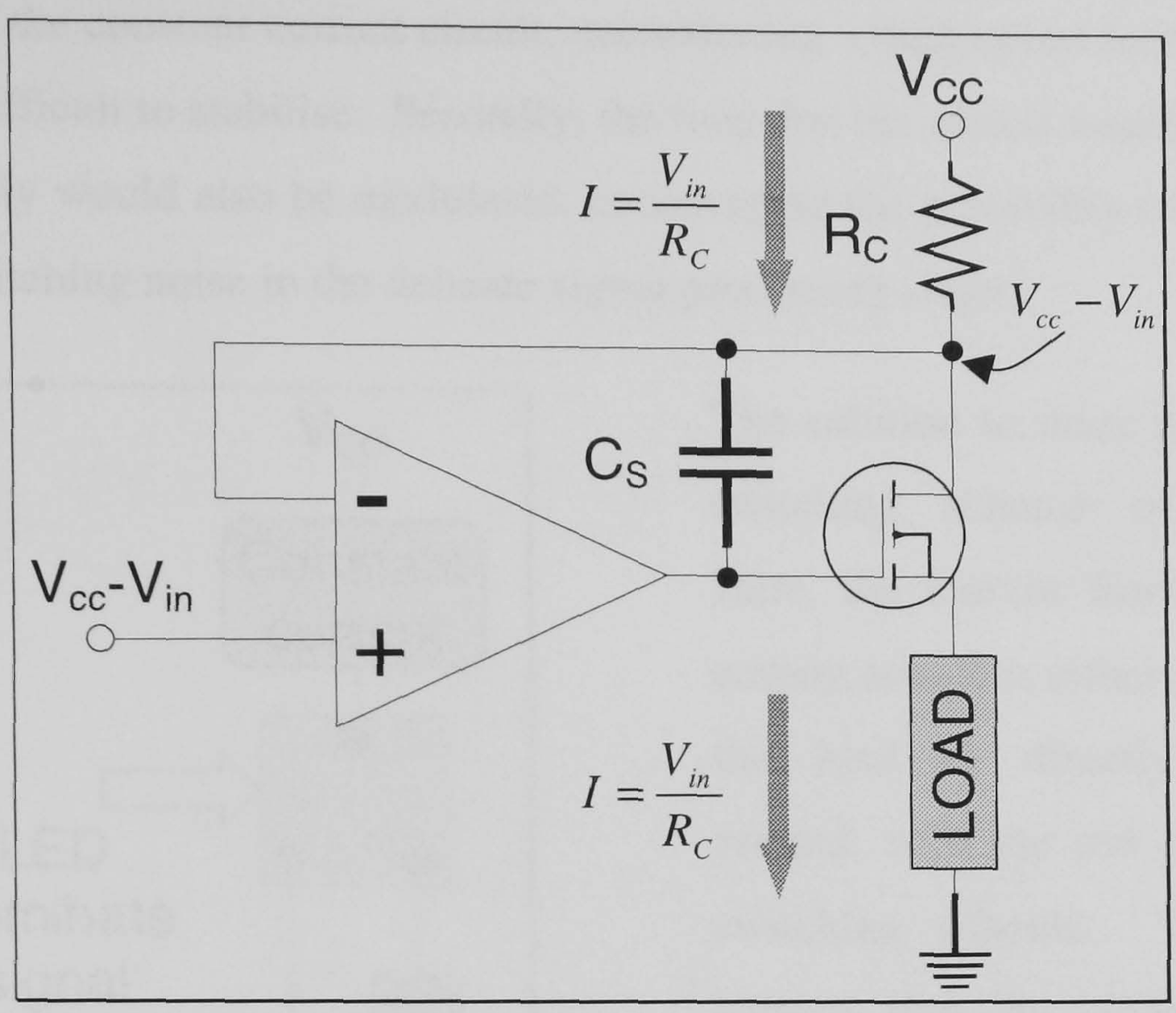


Figure I.21 - Current control methodology

Analysis of figure I.21 demonstrates the principle of operation. The supply voltage, V_{CC}, is measured and a control voltage, V_{in}, is subtracted from it and connected to the non-inverting input of the op-amp. The feedback action conspires to produce the exact drive voltage for the MOSFET transistor (P-channel) so that sufficient current flows in the resistor R_C to keep the op-amp input terminals at the same voltage; in other words to maintain a voltage of V_{in} across R_C. Thus, provided that the voltage drop across the load remains less than $V_{CC} - V_{in}$, a current of V_{in}/R_C will flow throught the load, independently of any variations in the load impedance (or voltage drop in the case of an LED). The bandwidth limiting capacitor, C_S, is included to maintain stability of the feedback loop and prevent both oscillations and overshoot in the step response.

I.3.1b Switching methodology

Although modulation of the LED current could be achieved by direct control of the input voltage to the constant current source of figure I.21, this would have several important disadvantages. Firstly, the modulation frequency would be limited by the

rise-time of the constant current circuit, necessitating a high-speed feedback loop that would be difficult to stabilise. Secondly, the load that the circuit would present to the power supply would also be modulated, encouraging the possibility of power-supply coupled switching noise in the delicate signal processing stages.

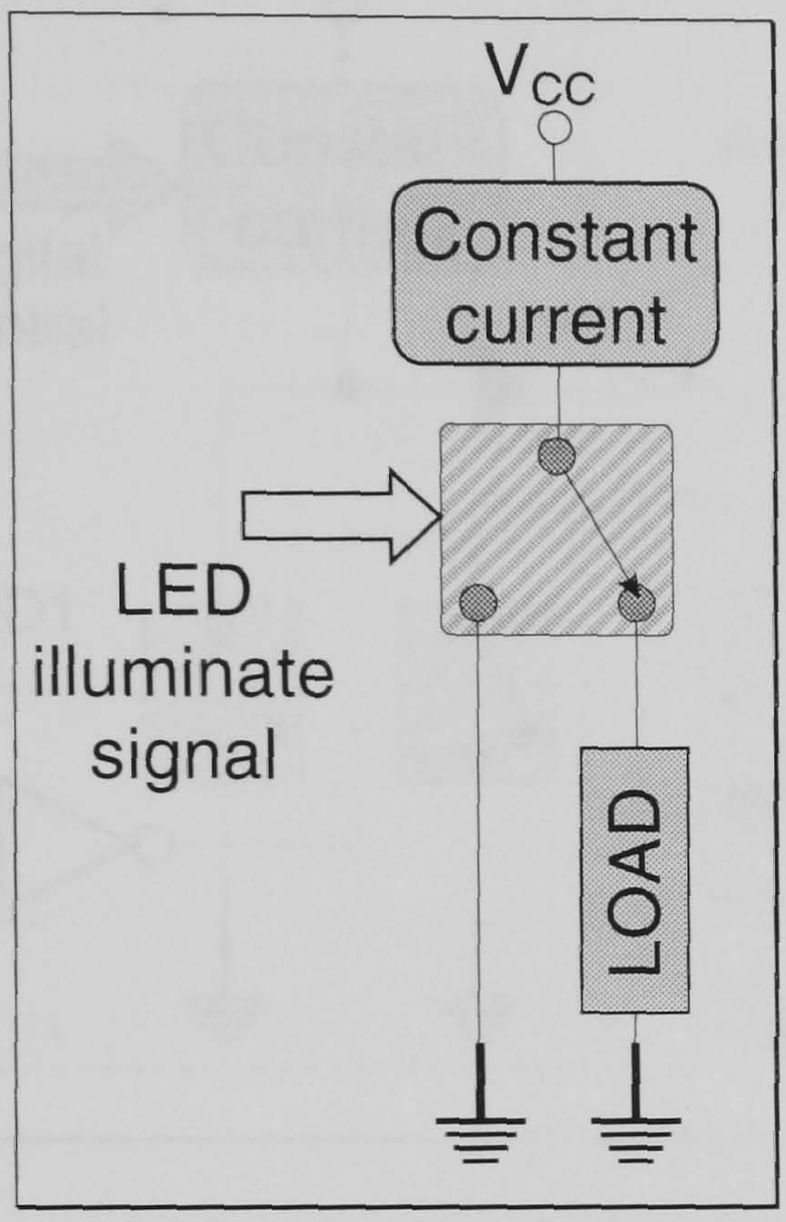


Figure I.22 - Switching methodology

The solution to these problems is the switching scheme of figure I.22. Here, the current from the constant-current source is either passed through the load or directly switched to ground, with the use of a compound switching scheme. This not only ensures that the circuit presents an invariant load to the power supply, but also implies that the maximum modulation frequency is determined only by the speed of the switches. It is now possible to use a stable and relatively low speed constant-current feedback-loop, whilst achieving a high modulation frequency. A pair of simple transistor-based ON-OFF switches that are driven in anti-phase can be used to implement the compound switching scheme (see figure I.23).

I.3.1c LED connections

Although it is important to be able to drive a pair of LEDs separately and independently, many commercial pulse oximetry probes place the LEDs in a back-to-back configuration (see figure I.23) to reduce the number of electrical connections to the probe. In that case the current polarity through the back-to-back connection determines which LED is illuminated, and therefore introduces the requirement that the circuitry must independently modulate and control current of both polarities. In fact, a practical realisation of the switching methodology shown in figure I.22 results

in circuitry that can, with the addition of only a single diode, be used in pairs to implement independent bipolar current control.

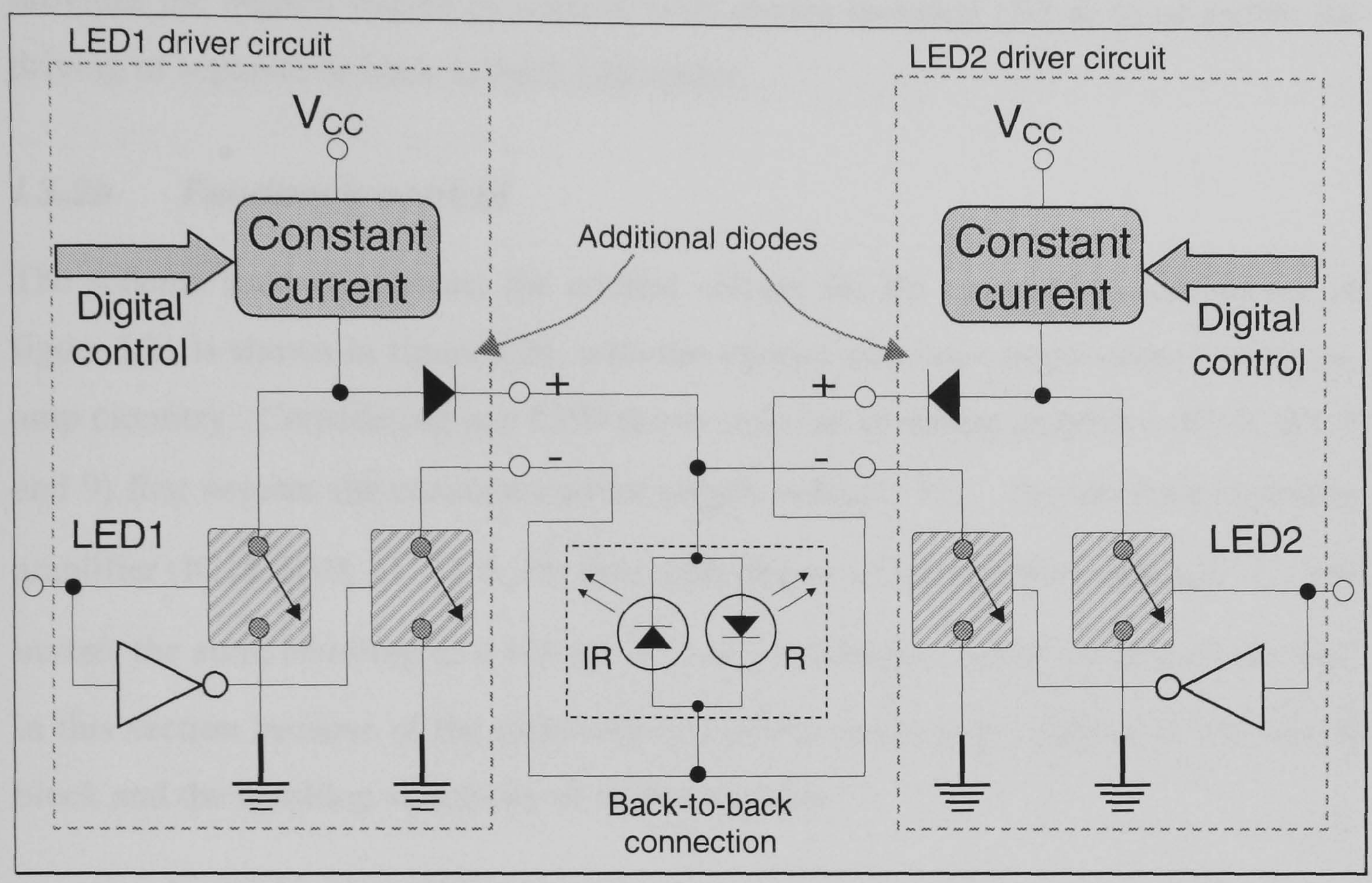


Figure I.23 - A pair of circuits driving a back-to-back LED connection

Figure I.23 shows a pair of LED driver circuits, with their additional diodes, driving a back-to-back LED connection. Because both LEDs cannot be illuminated simultaneously, and because the driver switches are operated in anti-phase, the diodes serve to isolate the circuits. Each driver circuit has separate and independent control of one polarity of current and therefore of one LED. Thus, the additional diodes enable the LED driver circuits to be used with either separate LEDs or a back-to-back connection without modification and with identical performance. This provides an increased degree of flexibility in the range of probes that can be driven by the circuitry.

I.3.2 Implementation

Implementation of a pair of constant current sources, building upon the methodology of figure I.21, is achieved using straightforward op-amp based circuitry. A digital

interface is provided that enables 8-bit (256 levels) control of the current supplied to each LED, up to a maximum of 100mA. Independent control of each current source provides the highest degree of control, with diodes included (D2 & 6) to enable the driving of separate or back-to-back LED pairs.

I.3.2a Feedback control

The scheme used to generate the control voltage for the constant current source of figure I.21 is shown in figure I.24, with the various functions implemented using op-amp circuitry. Considering one LED driver only, an inverting amplifier (IC18, R7, 8 and 9) first negates the constant-current supply voltage, V_{CC}. An inverting summing amplifier (IC18, R10, 11, 12 & 13) then adds this to an input control voltage, V_{in}, and inverts the sum, resulting in a voltage $V_{cc} - V_{in}$ as desired. Quad op-amps were used in this section because of the relatively low performance requirements of this circuit block and the resulting simplicity of implementation.

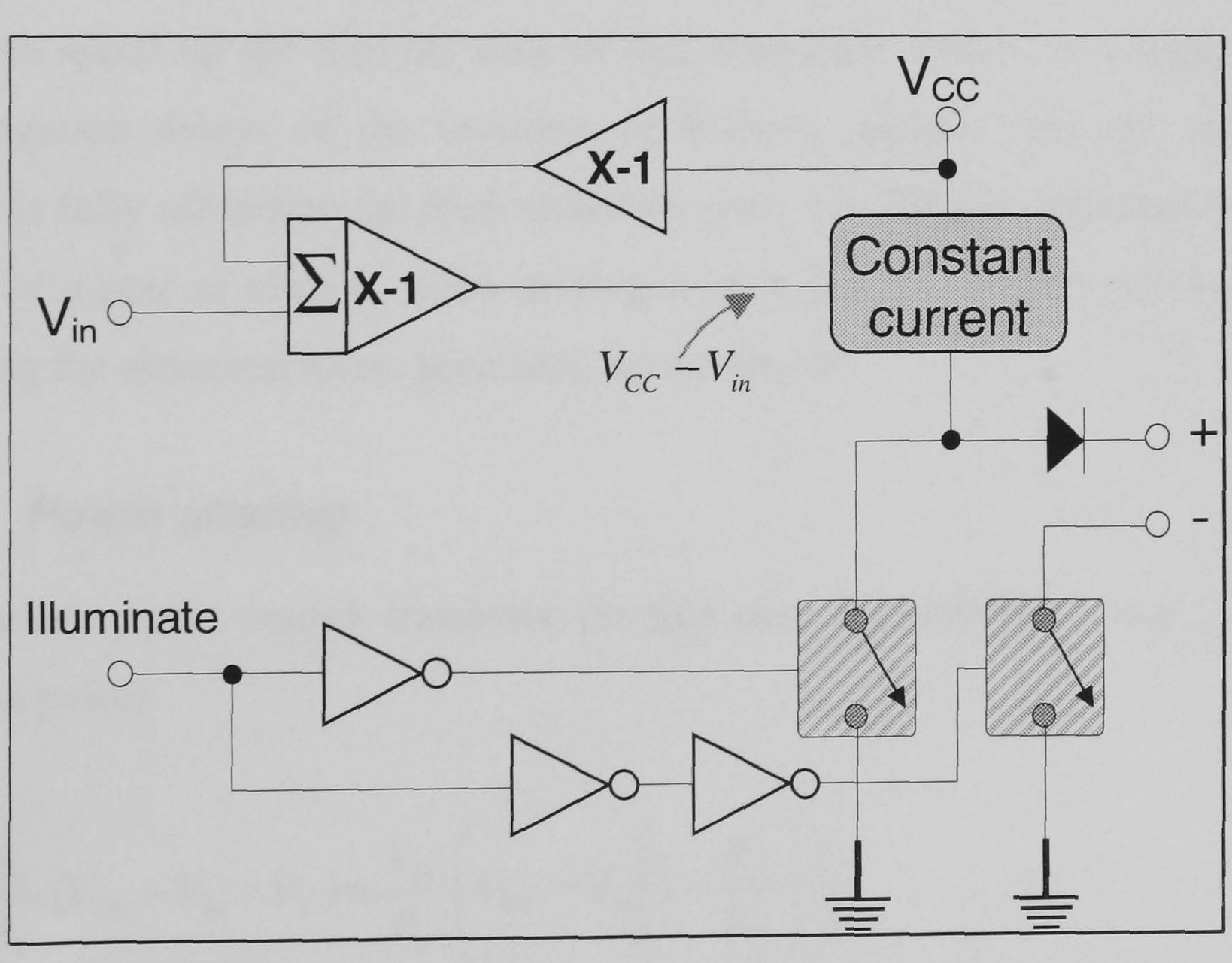


Figure I.24 - Constant current source implementation

A value of 50Ω was used for the current determining resistor, R_C (e.g. R14), defining the input voltage range as 0-5V for a current of 0-100mA (see section

I.3.1a). A bandwidth limiting capacitor of $100nF$ (e.g. C12) ensures stability of the constant-current feedback loop, whilst maintaining adequate frequency response. The measured current obtained using this circuitry is highly stable, independent of both supply voltage variations and changes in load impedance (or voltage drop), and linear with respect to the input voltage.

I.3.2b Switches

Anti-phase control of the switches is achieved using a logic inverter I.C. (IC13), which provides six individual inverters in a single package. Examination of figure I.24 highlights how three inverters can be utilised per driver circuit, resulting in homogenous buffers for both switches and therefore more predictable switch timing. NPN power transistors (e.g. TR2 & 3) are used to perform the switching, with base resistors (e.g. R19 & 21) determining the base current from the inverter outputs (5mA). Examination of the actual circuitry (figure I.20) shows an additional resistor (e.g. R20) connected from the base of one of the switches to ground. This resistor is included to speed up the turn-off time of that transistor, which, in conjunction with the propagation delays of the inverters, effectively ensures that the dummy-load transistor is fully off before the load transistor turns on. This is important in ensuring isolation of a pair of circuits when driving a back-to-back LED connection, and in minimising the electrical noise generated by the circuit.

I.3.2c Power sharing

The constant-current source transistor (in this case a MOSFET – e.g. TR1) must dissipate a power

$$P = \frac{V_{in}}{R_C}\left(V_{CC} - V_{in} - V_L\right) \equiv \frac{V_{in}}{R_C}\left(V_{CC} - V_{in}\left[1 + \frac{R_L}{R_C}\right]\right), \qquad \textbf{[I.9]}$$

where V_L is the LED forward voltage, or in the case of a linear load, R_L is the load impedance. Since equation [I.9] demonstrates that the power dissipation increases with the supply voltage for any particular load, it is therefore desirable to utilise the

minimum supply voltage necessary. Although a smaller supply voltage will restrict applicability of the circuit to smaller possible LED forward voltages, the expected range is relatively well defined. The maximum power dissipation by the transistor for any particular supply voltage can be found from equation [I.9],

$$P_{MAX} = \frac{V_{CC}^2}{4R_C}, \qquad \textbf{[I.10]}$$

occurring when $V_{in} = V_{CC}/2$ and with the minimum LED voltage drop, $V_L = 0$. If this circuit were directly supplied from the +12V overall supply then the maximum input voltage of 5V would be less than $V_{CC}/2$ and maximum dissipation would occur at $V_{in} = 5V$, resulting in a maximum power of 700mW. In the practical implementation a regulated supply (IC16 and associated circuitry) is used to provide a constant-current supply voltage of 9V. This reduced supply voltage results in the maximum power dissipation given by equation [I.10] occurring at $V_{in} = 4.5V$, with a maximum power of just 405mW; a maximum of 300mW is then dissipated by the voltage regulator. This power sharing methodology is important for surface-mount circuit layout, where restriction of the power dissipation in any one device eliminates the need for heatsinks or bulky device packaging. The use of a 9V supply defines a maximum LED voltage drop of 4V, high enough for the majority of practical devices.

I.3.2d Digital interface

The 0-5V control voltages are generated from an 8-bit digital word with the use of DAC-08 digital-to-analog converters (IC19 & 15). These devices are current-output DACs, and so transimpedance amplifiers (e.g. IC18 & R18) are used to convert the current into an output voltage. A stable 5V reference (e.g. IC21) is used to generate the reference current (0.55mA), resulting in a maximum output current of 1.1mA. A transimpedance gain of 4700 then converts this to a 0-5V output-voltage, which acts as an input to the constant current circuitry.

I.4 EQUALISATION CIRCUITRY

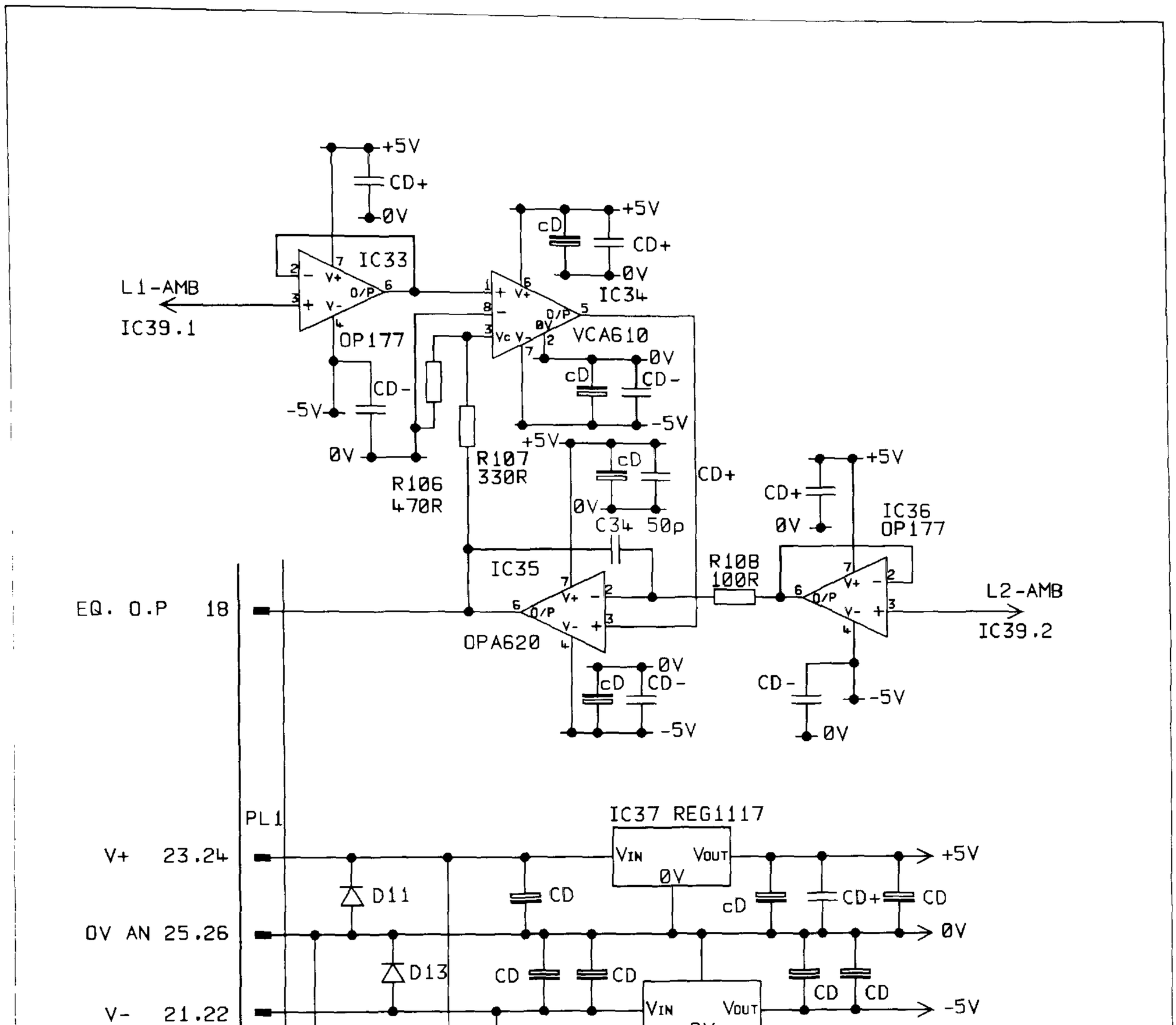


Figure I.25 – Schematic of equalisation circuitry

The design methodology for the equalisation circuitry has been discussed in detail in section 3.3, and so will not be repeated here. A practical implementation of the analogue circuitry described in section 3.3.4 is presented in figure I.25, and the following section merely describes the operation of this circuitry.

I.4.1 Principle of operation

The equalisation circuitry is based upon a voltage controlled amplifier, the VCA610 (IC34), which exhibits a power law characteristic between the gain, G, and an input control voltage, V_C,

$$G = 10^{-2(V_C+1)}. \qquad \textbf{[I.11]}$$

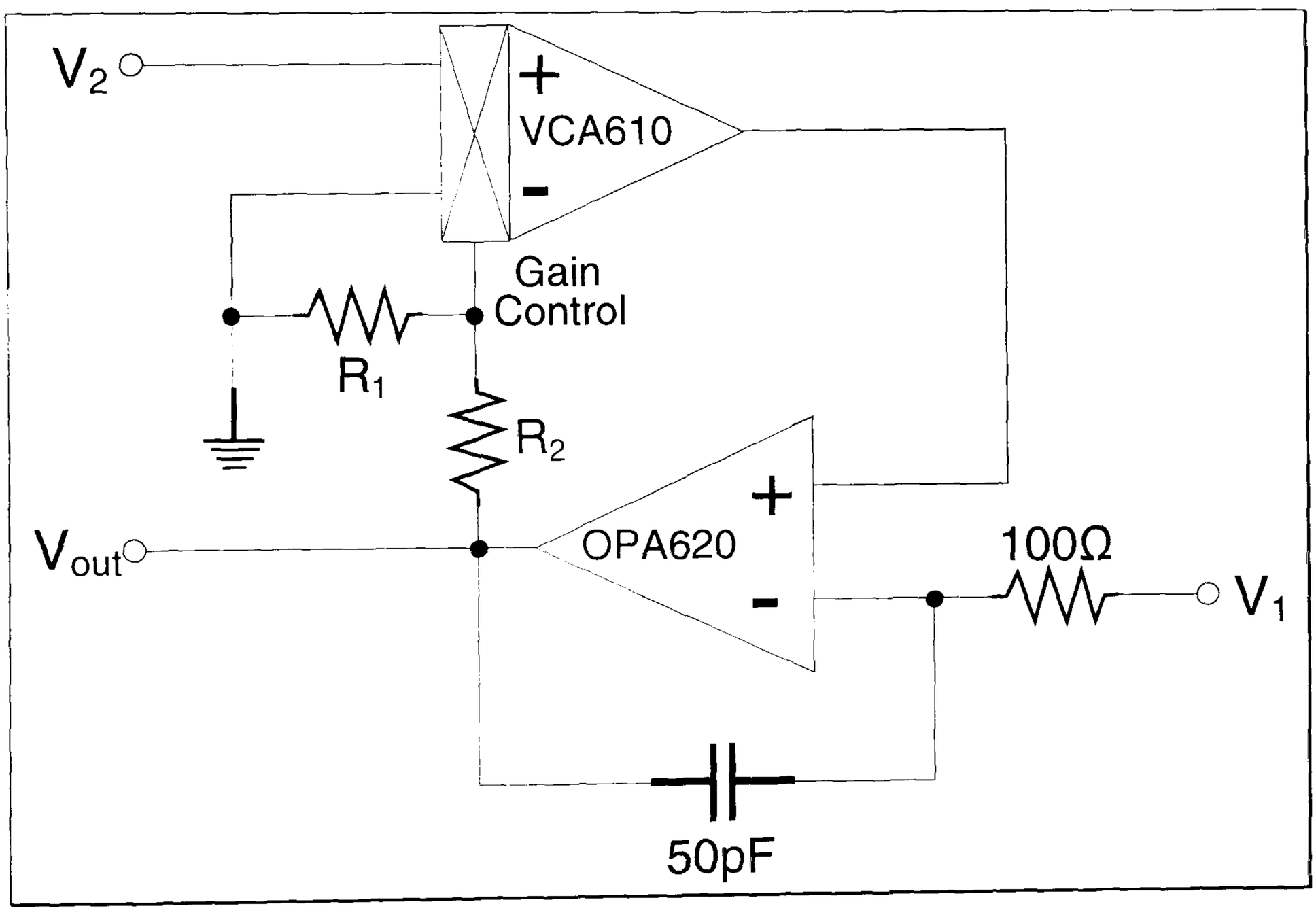


Figure I.26 – Logarithmic ratio circuitry

This amplifier is used as the non-linear element in the methodology of figure 3.6, resulting in the simplified implementation of figure I.26. Here an OPA620 (IC35) op-amp has been used in the feedback loop because of the high bandwidth requirement.

If the gain of this op-amp is A_V, then

$$V_{out}(t) = A_V\left(GV_2(t) - V_1(t)\right), \qquad \textbf{[I.12]}$$

where the voltage controlled amplifier gain is given by

$$G = 10^{-2\left(1+\frac{V_{out}(t)R_1}{R_2}\right)}. \qquad \textbf{[I.13]}$$

Simultaneous solution of equations [I.12] and [I.13] yields

$$V_{out}(t) = -\frac{R_1+R_2}{R_1}\left[1+\frac{1}{2}\log\left(\frac{\frac{V_{out}(t)}{A_V}+V_1(t)}{V_2(t)}\right)\right]. \qquad \textbf{[I.14]}$$

Equation [I.14] can be simplified by noting that $A_V \to \infty$, and rewriting in terms of natural logarithms,

$$V_{out}(t) \approx -\frac{R_1+R_2}{R_1}\left[1+\frac{1}{2\ln(10)}\ln\left(\frac{V_1(t)}{V_2(t)}\right)\right]. \qquad \textbf{[I.15]}$$

Equation [I.15] demonstrates that the output will contain a DC component that is a function of the equalisation circuitry, and an AC component that is proportional to the logarithmic ratio of the inputs, as desired. Because analysis of an equalised PPG signal only requires information about the signal dynamics (see section 3.1.1), the DC component introduced by the equalisation circuitry does not need to be considered.

I.4.2 Implementation details

The practical component values shown in figure I.25 result in output dynamics given by

$$V_{out}^{AC}(t) \approx -0.37 \ln\left(\frac{V_1(t)}{V_2(t)}\right), \qquad \textbf{[I.16]}$$

implying that an additional 8dB of gain should be applied to the equalised channel to produce dynamics of comparable size to those predicted theoretically.

Examination of figure I.26 also reveals an additional $50pF$ capacitor between the output and the inverting input of the op-amp. This capacitor ensures stability of the feedback loop by effectively reducing the gain of the op-amp to unity at high frequencies.

I.4.3 Power supply

The VCA610 operates from split ±5V supplies, necessitating some local regulation of the overall ±12V circuit supply. Although +5V supplies are generated elsewhere in the circuit, the use of local regulation (IC37 & 37 and associated circuitry) minimises interference from power-supply coupled noise. In addition, the high frequency of operation of the feedback loop requires good power-supply regulation to maintain stability, resulting in the abundance of decoupling capacitors seen in figure I.25. OP177 based followers (IC33 & 36) buffer the circuit inputs and limit them to a maximum excursion of ±5V, thus preventing the circuit inputs from exceeding the local power supplies.