



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

Research Commons

<http://waikato.researchgateway.ac.nz/>

## Research Commons at the University of Waikato

### Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

# Source Modulated Multiplexed Hyperspectral Imaging: Theory, Hardware and Application

A thesis submitted in fulfilment  
of the requirements for the degree of

**Doctor of Philosophy**

University of Waikato

by

**Lee Streeter**



THE UNIVERSITY OF  
**WAIKATO**  
*Te Whare Wānanga o Waikato*

2009



# Abstract

The design, analysis and application of a multiplexing hyperspectral imager is presented. The hyperspectral imager consists of a broadband digital light projector that uses a digital micromirror array as the optical engine to project light patterns onto a sample object. A single point spectrometer measures light that is reflected from the sample. Multiplexing patterns encode the spectral response from the sample, where each spectrum taken is the sum of a set of spectral responses from a number of pixels. Decoding in software recovers the spectral response of each pixel. A technique, which we call complement encoding, is introduced for the removal of background light effects. Complement encoding requires the use of multiplexing matrices with positive and negative entries.

The theory of multiplexing using the Hadamard matrices is developed. Results from prior art are incorporated into a singular notational system under which the different Hadamard matrices are compared with each other and with acquisition of data without multiplexing (pointwise acquisition). The link between Hadamard matrices with strongly regular graphs is extended to incorporate all three types of Hadamard matrices. The effect of the number of measurements used in compressed sensing on measurement precision is derived by inference using results concerning the eigenvalues of large random matrices. The literature shows that more measurements increases accuracy of reconstruction. In contrast we find that more measurement reduces precision, so there is a tradeoff between precision and accuracy. The effect of error in the reference on the Wilcoxon statistic is derived. Reference error reduces the estimate of the Wilcoxon, however given an estimate of the Wilcoxon and the proportion of error in the reference, we show that Wilcoxon without error can be estimated.

Imaging of simple objects and signal to noise ratio (SNR) experiments are used to test the hyperspectral imager. The simple objects allow us to see that the imager produces sensible spectra. The experiments involve looking at the SNR itself and the SNR boost, that is ratio of the SNR from multiplexing to the SNR from pointwise acquisition. The SNR boost varies dramatically across the spectral domain from 3 to the theoretical maximum of 16. The range of boost values is due to the relative Poisson to additive noise variance changing over the spectral domain, an effect that is due to the light bulb output and detector sensitivity not being flat over the spectral domain. It is shown that the SNR

boost is least where the SNR is high and is greatest where the SNR is least, so the boost is provided where it is needed most. The varying SNR boost is interpreted as a preferential boost, that is useful when the dominant noise source is indeterminate or varying. Compressed sensing precision is compared with the accuracy in reconstruction and with the precision in Hadamard multiplexing. A tradeoff is observed between accuracy and precision as the number of measurements increases. Generally Hadamard multiplexing is found to be superior to compressed sensing, but compressed sensing is considered suitable when shortened data acquisition time is important and poorer data quality is acceptable.

To further show the use of the hyperspectral imager, volumetric mapping and analysis of beef *m. longissimus dorsi* are performed. Hyperspectral images are taken of successive slices down the length of the muscle. Classification of the spectra according to visible content as lean or nonlean is trialled, resulting in a Wilcoxon value greater than 0.95, indicating very strong classification power. Analysis of the variation in the spectra down the length of the muscles is performed using variography. The variation in spectra of a muscle is small but increases with distance, and there is a periodic effect possibly due to water seepage from where connective tissue is removed from the meat while cutting from the carcass. The spectra are compared to parameters concerning the rate and value of meat bloom (change of colour post slicing), pH and tenderometry reading (shear force). Mixed results for prediction of blooming parameters are obtained, pH shows strong correlation ( $R^2 = 0.797$ ) with the spectral band 598–949 nm despite the narrow range of pH readings obtained. A likewise narrow range of tenderometry readings resulted in no useful correlation with the spectra.

Overall the spatial multiplexed imaging with a DMA based light modulation is successful. The theoretical analysis of multiplexing gives a general description of the system performance, particularly for multiplexing with the Hadamard matrices. Experiments show that the Hadamard multiplexing technique improves the SNR of spectra taken over pointwise imaging. Aspects of the theoretical analysis are demonstrated. Hyperspectral images are acquired and analysed that demonstrate that the spectra acquired are sensible and useful.

# Acknowledgements

L. Streeter acknowledges the support of an New Zealand Tertiary Education Commission Enterprise Doctoral Scholarship, ID UOWX05001 and a University of Waikato Doctoral Extension Scholarship. The NIR imaging project is part of the meat quality project, a New Zealand Foundation for Research Science and Technology funded AgResearch initiative, ID FRST8807.

Technical and academic support is acknowledged. The advice on statistics by Martin Upsdell and meat science by Carrick Devine are acknowledged, as is the expert meat butchering, handling and advice of Kevin Taukiri and Peter Dobbie. The fabrication of electronic equipment and support by Keith Stantial and Stuart Finlay and fabrication of other hardware by Brian Atkins and Keith Hill are acknowledged. The computer technical support at UoW of Bruce Rhodes and David Nichols and optics technical support of Pawan Shrestha are gratefully appreciated.

## Personal Acknowledgements

My appreciation goes to my supervisors Michael Cree, Rob Burling-Claridge and Rainer Künnemeyer for seeing me through my studies with advice and direction. I would like to thank my wife Emma for the constant support and encouragement. I would also like to thank my parents Tonie and Jennifer for always encouraging me in my academic endeavours, and indeed any endeavour I took up.

For my late Father.  
Tonie Streeter 1951–2007



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Multiplexing . . . . .	3
1.2 Theory . . . . .	4
1.3 Hardware and Experimental Direction . . . . .	6
1.4 Thesis Structure . . . . .	7
1.5 List of Publications . . . . .	8
1.5.1 Journal Publications . . . . .	8
1.5.2 Conference Papers . . . . .	9
<b>2 Review</b>	<b>11</b>
2.1 Multiplexing . . . . .	11
2.2 Hardware . . . . .	12
2.3 Hadamard Transform . . . . .	16
2.4 Compressed Sensing . . . . .	20
2.5 Near Infrared Spectroscopy and Chemometrics . . . . .	23
2.5.1 PCR and PLS . . . . .	24
2.5.2 Spectral Preprocessing . . . . .	26
2.5.3 Extended Canonical Variates Analysis . . . . .	27
2.6 Data Analysis Methods . . . . .	28
2.6.1 Correlation . . . . .	28
2.6.2 Variography . . . . .	29
2.6.3 Receiver Operator Characteristics and the Wilcoxon . . . . .	29
2.7 Near Infrared Spectroscopy and Beef . . . . .	30
<b>3 Theory</b>	<b>37</b>
3.1 Multiplexing . . . . .	38
3.1.1 Hadamard Matrices . . . . .	38
3.2 Error Effects Under Hadamard Encoding . . . . .	40
3.2.1 Additive Random Noise . . . . .	41
3.2.2 Additive Offset Bias . . . . .	43
3.2.3 Poisson Noise . . . . .	44
3.2.4 Multiplicative Drift, Systematic Multiplicative Bias . . . . .	47
3.2.5 Combined Effect . . . . .	49

3.3	Graph Theory and Hadamard Multiplexing . . . . .	51
3.3.1	The ‘Link’ Between $W$ and $S$ , $G$ and $H$ . . . . .	52
3.4	Compressed Sensing . . . . .	56
3.4.1	CS and the Effect of Random Noise . . . . .	58
3.5	The Effect of Error in the Reference on the Wilcoxon . . . . .	60
3.5.1	Alternative Derivation that the Wilcoxon is Equivalent to the AUC . . . . .	61
3.5.2	The Effect of Error in the Reference . . . . .	62
3.5.3	Numerical Simulation . . . . .	64
<b>4</b>	<b>Hardware</b>	<b>67</b>
4.1	Optics and Operational Theory . . . . .	68
4.1.1	Data Acquisition . . . . .	71
4.1.2	Complement Encoding . . . . .	72
4.1.3	Decoding and Noise in Complement Encoding . . . . .	75
4.1.4	Reference Beam Correction . . . . .	77
4.1.5	Reference Object Correction: the White Tile Effect . . . . .	79
4.2	Testing Methodology . . . . .	81
4.2.1	G-matrix (Complement S-matrix) Multiplexing . . . . .	82
4.2.2	Compressed Sensing . . . . .	83
4.2.3	Objects Imaged . . . . .	84
4.3	Results and Discussion . . . . .	87
4.3.1	Images and Spectra . . . . .	87
4.3.2	Signal to Noise Ratios and Noise Features . . . . .	94
4.3.3	Comparison of Compressed Sensing with Hadamard Imaging . . . . .	98
<b>5</b>	<b>Volumetric Mapping of Beef <i>M. Longissimus Dorsi</i></b>	<b>103</b>
5.1	Meat Collection and Instrumental Apparatus . . . . .	104
5.1.1	Meat Collection . . . . .	104
5.1.2	Instrumental Configuration . . . . .	105
5.1.3	Data Collection . . . . .	105
5.1.4	Data Summary . . . . .	109
5.2	Lean Pixel Classification . . . . .	110
5.3	Variography . . . . .	117
5.4	Bloom . . . . .	120
5.5	Tenderometry and pH Correlation . . . . .	122
5.5.1	pH . . . . .	122
5.5.2	Tenderometry . . . . .	125
<b>6</b>	<b>Conclusion and Future Work</b>	<b>127</b>
6.1	Theory . . . . .	128
6.2	Hardware . . . . .	130
6.3	Application . . . . .	133
6.4	Future Directions . . . . .	135
	<b>References</b>	<b>139</b>

# List of Figures

2.1	Simplified diagram of a spectrometer. The arrows indicate the light path. Light entering through the entrance slit is dispersed and reflected onto the sensor. . . . .	13
2.2	Diagram of a multiplexing spectrometer. The arrows indicate the light path. Spectral wavelength components separated by the dispersion element are selectively passed and occluded by the encoder and measured by the sensor. Successive measurements utilise different sets of wavelengths, encoding the spectrum. . . . .	14
2.3	DMA and CCD based imaging spectrometer. The DMA modulates the spectral and/or spatial information which is then recorded by the CCD. . . . .	15
3.1	SNR of CS for a numerical example and the SNR due to the theoretically inferred precision. As the number of measurements increases the random error dominates the SNR. . . . .	59
3.2	Numerical computation and theoretical prediction of the Wilcoxon with varying amount of error in the reference. . . . .	65
4.1	(a) A DMA chip and (b) a close up drawing of the mirrors. (Source: <a href="http://www.dlp.com/tech/what.aspx">http://www.dlp.com/tech/what.aspx</a> , date accessed 16/02/2009, ©Copyright 2009 Texas Instruments Incorporated. Used by permission.) . . . . .	68
4.2	Diagram of the optical setup. Arrows indicate the light path. The spectrometer measures light from the sample and light piped directly from the source. The light chopper rate is controlled by the spectrometer. . . . .	68
4.3	a) The optical arrangement. b) A close up of the projector without the reference beam fibre. . . . .	69
4.4	Illustration of splitting the Hadamard matrix into $H^+$ and $H^-$ components. In $H^+$ the $-1$ s of the original H-matrix are converted to 0's. In $H^-$ the $+1$ s are converted to 0s and the $-1$ 's are converted to $+1$ s. . . . .	73
4.5	The testpattern for examining spatial variation in the reconstruction. The rate of change of the bright-dark patterns increases from the top left to the bottom right. . . . .	85
4.6	The wood and acrylic object. The acrylic is on the left and the wood on the right. . . . .	85
4.7	The wood and polystyrene object. . . . .	88
4.8	The acrylic and PCB object. The acrylic is on the left and the PCB covering the acrylic on the right. Marks are imperfections such as scuffs on the material. . . . .	88

4.9	Images of the test pattern. a) A $16 \times 16$ pixel image and b) a $64 \times 64$ pixel image. . . . .	89
4.10	Acrylic and wood object imaged using G-matrix encoding. a) Reference corrected spectra and b) the standard normal variate transform of the same spectra. . . . .	90
4.11	The first principal component of the image of the wood and acrylic object.	90
4.12	Wood and polystyrene object imaged using G-matrix encoding. a) Reference corrected spectra and b) the standard normal variate transform of the same spectra. . . . .	91
4.13	The second principal component of the image of the wood and polystyrene object. The imaging was performed with a 255 order G-matrix. . . . .	92
4.14	Spectra from the eighth row of an acrylic and white cardboard object. . .	92
4.15	The thirtieth spectral image of the acrylic and cardboard object at 1150 nm. 93	
4.16	Spectra acquired by H-matrix encoding. The wood and polystyrene object is on the left and the acrylic and PCB fibreboard object is on the right. . .	94
4.17	a) The wood and polystyrene object at 1533 nm imaged using H-matrix encoding. b) The PCB and fibreglass object. . . . .	94
4.18	a) SNRs for the G-matrix encoding with no reference correction, correction per acquired spectrum and the average reference spectrum over an acquisition period for the entire encoding matrix. b) Significance test of reference correction using the of the reference measurements and the per sample reference measurements. . . . .	95
4.19	The SNR boost over pointwise imaging due to G-matrix encoding. . . . .	96
4.20	The average SNR due to H-matrix encoding. . . . .	96
4.21	The SNR boost over pointwise imaging due to H-matrix encoding. . . . .	97
4.22	The ratio of the H-matrix SNR boost to the G-matrix SNR boost. . . . .	98
4.23	The average concatenated rows from Hadamard imaging of the Spectralon tile. This represents the best estimate of the true pixel values. . . . .	99
4.24	Signal to noise ratios for Hadamard, compressed sensing and pointwise imaging. . . . .	100
4.25	SNR boost comparison between Hadamard encoding and CS. . . . .	101
4.26	SNR against number of measurements for CS. The Hadamard and pointwise lines are shown purely for reference and do not imply Hadamard or pointwise imaging with fewer than 256 measurements. . . . .	102
4.27	Relative error in CS reconstruction against the number of measurements. .	102
5.1	Diagram of the experimental setup. . . . .	105
5.2	Illustration of the action of the morphological reconstruction. . . . .	106
5.3	Images of animal 5, left hand side, slice eight. . . . .	108
5.4	Average spectra from the slice images. . . . .	111
5.5	The classification vector. . . . .	112
5.6	Box and whisker plot of the first score per animal. . . . .	112
5.7	Box and whisker plot of the second score per animal. . . . .	113
5.8	Box and whisker plot of the third score per animal. . . . .	113
5.9	Box and whisker plot of the overall classifier output per animal. . . . .	114
5.10	ROC curve for the validation data. . . . .	115

---

5.11	The average variogram for each wavelength down the length of the LDs. Each line represents one wavelength. The variograms for all the wavelengths are shown together in this manner so that the general shape of the variation against lag down the muscle can be seen. . . . .	117
5.12	The average variogram for each wavelength down the length of the LDs after EMSC. The effect due to water and the average lean spectrum was removed before computation of the variograms. . . . .	118
5.13	Example bloom curves from animal 6, left side, and the parametric curve fit to them. . . . .	120
5.14	The pH readings down twelve LD muscles (the first animal is excluded). Vertical dashed lines separate one muscle from the next. The data are ordered per animal, left side followed by the right side. There is a tendency for the right side pH readings to be slightly higher than the left. . . . .	123
5.15	Correlation (r value) of each wavelength with the pH. The average of the selected region indicated is compared with the pH reference values. . . .	123
5.16	Scatter of the mean NIR reflectance in the band 598–949 nm against the pH.	124
5.17	Correlation (r value) of tenderness readings against the wavelengths. . .	125



# List of Tables

5.1	Enumeration of the breeds from which the muscles were taken. . . . .	109
5.2	Breakdown of the number of spectra per animal obtained after the removal of air spectra. . . . .	109
5.3	Summary of the reference data. . . . .	110
5.4	AUC(SE) for pixel classification. . . . .	115
5.5	Crossvalidation result of modelling bloom parameters by NIR spectra. . .	121



# Chapter 1

## Introduction

The purpose of this thesis is the design, construction, and analysis of a hyperspectral imager that operates in the near infrared (NIR) and if possible the visible (Vis) domain. A hyperspectral imager is an optical system that acquires images, where each pixel is comprised of a full spectrum. The spectrum at each pixel, while discrete, consist of narrow and closely spaced bands to approximate a continuous signal. A conventional spectrometer measures the spectral detail totalled over a given physical area on an object or sample. The purpose of hyperspectral imaging is to acquire detailed spectral information with spatial specificity, that is, a hyperspectral imager distinguishes the spectral detail from one location on an object to another.

In building a hyperspectral imager we need to characterise the system for salient features such as signal to noise ratio and related limitations with respect to image acquisition time and spatial resolution. In other words if we have a system that acquires hyperspectral images it is important to assess how well it does that job. We seek a complete and unified theoretical model for understanding the behaviour and performance of the imager. The theoretical model encompasses all forms of error that we can realistically expect to see in the system. Some forms of error are systematic and correctable, where others are random. The model indicates how correction of the systematic error forms will affect the overall performance of the imager.

At the outset of this research spectrometers that examine the spectral region from 400–1700 nm were available for use in this research. The spectrometers by default are configurable to examine a quarter inch diameter region through a fibre optic probe or to integrate light from an area tens of millimetres in diameter. Since we can spectrally analyse the sum total of light from a given region the goal is to somehow modulate the light so that we can obtain spectral information with spatial localisation.

Alternative methods of spectral image capture that involve two dimensional sensors, typically via the so called pushbroom and whiskbroom methods, can provide fast capture of detailed spatial and spectral information. Such methods are dismissed at the outset

as unsuitable for this research as two dimensional sensors that act beyond 1100 nm are expensive, and construction of a such a system essentially requires building new spectrometers which is unnecessary when good working spectrometers were available. Furthermore, in utilising a two dimensional sensor to perform hyperspectral imaging one typically uses the rows of the sensor as individual spectrometers that analyse different spatial locations in parallel. Different spectrometers have different behaviour in terms of both random noise and systematic error. Analysis and models built for the data produced may not uniformly fit to the spectra from the different rows of the sensor, causing error. The process of correcting such issues is known as calibration transfer, a process that we choose to avoid.

Having made the decision to utilise a single point spectrometer to perform hyperspectral imaging we now decide on what optical system to use to obtain spatial information. The options fall into two broad classes: acquiring spectra one pixel at a time or some method of parallel acquisition such as multiplexing. Modern light modulation devices (discussed below) provide the flexibility to do both pointwise and multiplexed data acquisition with the same hardware. Furthermore, the noise reduction properties of multiplexing makes a multiplexed system the natural choice.

The system requires a light source, control of the light path, and integration and spectroscopic analysis of the light from the sample. The methods of light modulation broadly fall into two classes: those that control the light path to the sample being imaged (which we call source modulation) and those that control the light path from the sample to the sensor (call this sample modulation). A source modulation system forms patterns on the sample so that only certain parts of the sample are illuminated, where different areas of the sample are sequentially illuminated in turn. In contrast, a sample modulation system floods the sample with light and controls which areas of the sample are permitted to illuminate the sensor. A system that does both is conceivable but for our purposes is unnecessarily complex. In Section 3.2.2 we see theoretically that modulation of light after the sample introduces undesirable issues surrounding background offset correction, thus a source modulation system is preferable. Specifically considering our system, multiplexing both the spatial and spectral domains is conceivable, however because of the availability of good working spectrometers spectral domain multiplexing is not necessary. In summary, we seek to design and build a hyperspectral imaging system comprised of a light source modulation system that projects multiplexing patterns on to the sample and a spectrometer that integrates and spectrally analyses the light from the sample.

Modulation of light to the sample is most easily and practicably achieved by using an optical image projector. The optics for light projection systems are well understood and easy to assemble, and are present in common slide projectors and digital data projectors. Slide projectors use tungsten halogen bulbs, the same type of bulb that is commonly used

in visible/near infrared spectroscopy, whereas modern projectors use mercury arc lamps or metal halide lamps which have undesirable spectral characteristics, so we choose a traditional slide projector arrangement. An optical component that forms the light patterns that encode the pixels is required as part of the optical pathway. We desire flexible computer control of the light patterns projected onto the sample so clearly traditional slides are not acceptable. Options for achieving digital control of pattern formation include use of a liquid crystal device (LCD) or a digital micromirror array (DMA). LCDs are designed for visible domain light modulation and have strong spectral features whereas the broad spectral characteristics of mirrors make DMAs the natural choice. For these reasons we choose a traditional slide projector with a DMA in place of the slides as the light pattern generation device to comprise the light modulation system.

The analysis of the imager presented in this thesis involves characterisation of the performance with respect to random noise and systematic error, and evidence that the imager produces spectroscopic data of adequate quality for spectroscopic applications. The purpose of multiplexing is to reduce noise relative to per pixel acquisition. To understand the effect and behaviour of noise under multiplexing we construct a general model of multiplexing with all reasonable possible error sources. We seek a unified notation under which different mathematical encoding systems can be analysed and compared. We interpret the error sources in the context of light multiplexing and analyse the form of the error that we expect of our system. Informed by the model, signal to noise ratio experiments are conducted and interpreted. Spectroscopic images are taken of simple objects with interesting spectral shape to demonstrate that the system gives useful data.

To further strengthen the case that the imager produces useful data we turn to an application closer to a real world scenario. Spectral images are taken of beef strip loins to volumetrically map the spectral characteristics. The strip loin is the source of many important retail cuts and non destructive prediction of quality features by spectroscopy can potentially add value to the meat. In performing spectroscopic measurement an understanding of the source and form of spatial variability is informative. Furthermore, quality assessments must be computed on spectra of the lean and not of the fat or any other visible constituent. To this end we investigate the classification of spectra as lean or not lean.

## 1.1 Multiplexing

Multiplexing is the method of mixing multiple objects according to specific grouping patterns. We use multiplexing specifically for signal measurement. Consider weighing a number of objects that have an average weight of 10 g, but the only weighing device available has an uncertainty of 1 g, so that, on average, the signal to noise ratio of measurement is 10. Here ‘noise’ is identified with the uncertainty and is estimated as the

standard deviation in measurement. Now consider a multiplexing system which weighs groups of ten objects at a time, thus the average signal to noise ratio of the measurement of a group is 100. After measuring  $N$  different combinations, with appropriate selection of the combinations, then the weights of  $N$  different objects can be estimated. If the grouping patterns are well selected then the estimates of the object weights have less noise than measurement of each object individually.

In this thesis we focus primarily on Hadamard matrices as the grouping pattern. In particular, we primarily utilise the Hadamard H-matrix which is the matrix with maximum determinant when all entries have value between negative one and positive one. Other Hadamard matrices, namely the G-matrix and the S-matrix, which can be derived from the H-matrix are also given focus. The effect of Hadamard multiplexing on additive noise is well known (Harwit and Sloane, 1979), specifically, if an order  $N$  H-matrix is used then the noise variance is reduced by a factor of  $N$ .

In the context of optical light multiplexing the measurement device is a light sensor such as an array of light sensitive diodes in a spectrometer. For the type of hyperspectral imaging device described, multiplexing involves the selective illumination of pixels according to the multiplexing pattern matrix. A DMA based illumination system can be easily adapted to perform multiplexing with any conceivable encoding matrix without the need to change or adapt any hardware. Multiplexing light immediately causes issues surrounding light drift, stray and background light and noise that has variance proportional to the light intensity such as photon fluctuations.

## 1.2 Theory

The theoretical analysis presented herein (Chapter 3) is primarily concerned with the construction of a general model for multiplexing that accounts for all reasonable possible error sources. A uniform notation is developed under which the error sources are analysed for the three Hadamard matrices and other multiplexing systems. The error sources include systematic and random forms that both manifest as multiplicative and additive effects resulting in four main classes of error. The effect of multiplexing on random additive noise is well known, namely that for a general ‘good’ multiplexing matrix, for multiplexing  $N$  measurements, then the variance is reduced by a factor proportional to  $N$ . The effect of Hadamard multiplexing on random noise with variance proportional to the signal, i.e. Poisson noise, has been studied by others for the S-matrix and to a much lesser extent the H-matrix (Harwit and Sloane, 1979; Nitzsche and Riesenber, 2003; Hassler *et al.*, 2005; Damaschini, 1993). Poisson noise variance is increased by S-matrix multiplexing, whereas H-matrix multiplexing has no effect on average. The H-matrix analysis the prior art was developed for the specific context of time resolved fluorescence imaging

and considered each pixel as a separate emitter. In contrast, we treat the light source as the single emitter and consider the multiplexing apparatus and the sample imaged as attenuating the light. Systematic errors occur due to background light (additive) and drift in the light source (multiplicative). Systematic additive error is reduced by multiplexing in the same way random additive error is, but, by definition, systematic additive error does not have zero expected value at any given time point or else it does not exist. Multiplicative systematic error is mixed by the multiplexing so the effect is more complicated than additive error. Both the form of the error and the effect on the random noise due to correction of systematic error are explored.

Recently Ratner *et al.* (2007) linked a type of graph, called the strongly regular graph (SRG), to Hadamard multiplexing. Specifically the SRG was linked to Hadamard matrices with ones and zeros as entries, namely the S-matrices. The link was established via the respective eigenvalue structures and the observation that the effect on noise was computable from the trace, that is, the sum of the eigenvalues. Under certain conditions an SRG has the same eigenvalue structure as an S-matrix, however, the SRG has the distinct advantage over the S-matrix that the number of objects included in each measurement is variable. Reducing (or increasing) the number of objects included in each measurement means that certain error processes, such as Poisson noise or sensor saturation, are reduced in a superior manner than with Hadamard matrices. Ratner *et al.* (2007) do not consider multiplexing matrices that include negative one entries. We extend the theory to examine the link, if any, between SRGs and the other Hadamard matrices.

Compressed sensing (CS) (Donoho, 2006) is a recent technique for signal acquisition that falls in the class of multiplexing. Traditional theory tell us that if we want to acquire  $N$  data, then we need to take at least  $N$  measurements. Traditional multiplexing requires  $N$  measurements according to a specific linear transformation from which the data are recovered by inversion of the transformation. Compressed sensing theory tells us that if the signal is compressible then we can get a good reconstruction of  $N$  data from fewer than  $N$  measurements, i.e., inversion of an underdetermined system. Furthermore, matrices with random entries form good patterns for CS encoding. The literature shows that CS is accurate even when random noise is present. We examine by inference the effect on precision (as a distinct measure from accuracy) due to CS, estimated as the signal to noise ratio due to random error. The action of the number of measurements on the precision is analysed via the eigenvalue structure of large random matrices. In considering the accuracy of CS reconstruction it is recognised in the literature that a rectangular matrix of random entries resembles an orthonormal system to within certain limits, the so-called restricted isometry principle. We utilise a similar idea to examine by inference how the precision in CS is affected by the number of measurements.

The Wilcoxon and the area under the curve (AUC) of the receiver operator characteris-

tic (ROC) are two equivalent statistics for estimating the probability that one group of data have values that are larger than another group. Given data that falls into two groups, the Wilcoxon is the ratio of the count times that the members of one group are larger in value than the members of the other group, over the total number of comparisons. Similarly the AUC ROC is a measure of the separation between the two groups, see section 2.6.3. The Wilcoxon and AUC are important statistics in the data analysis in the application of the hyperspectral imager. To compute the Wilcoxon we need to know what group each datum is assigned to. To know the group assignments typically a reference method is used, however the reference method is a classification method that may (and probably does) have errors in the form of false positives and false negatives. We derive the Wilcoxon from the AUC, hence establishing the link directly, and then use the result to compute the effect of error in the reference on the Wilcoxon.

### 1.3 Hardware and Experimental Direction

There are two purposes for the experimental work presented in this thesis, the first is to characterise the error behaviour of the hyperspectral imager and the second is to demonstrate that the imager produces spectra of sufficient quality for use in an application. The characterisation of the error behaviour is performed via estimation of the signal to noise ratio (SNR) of the spectra and examination of the improvement in SNR when multiplexing over measuring one pixel at a time. SNR experiments are conducted for H-matrix and G-matrix multiplexing and compressed sensing. The Hadamard multiplexing is considered accurate so only precision is examined via the SNR. For compressed sensing both precision and accuracy are assessed, where accuracy is examined by considering the difference between the compressed sensing result and the best estimate of the true spectral pixel values. In our case the best possible estimate of the true pixel values is the average of several Hadamard hyperspectral images.

Correction of the systematic error and the effect of the corrections on random noise sources are necessary. Only one hardware configuration is tested, namely the DMA based source modulated multiplexed system. To correct for background illumination we devise the concept of complement encoding, in which the measurement of pixel responses is divided into two complement groups. Background illumination occurs in both the complementary encodings. If the complementary spectra are captured quickly enough then the background illumination is the same for both acquisitions and the encoding-decoding scheme removes the background lighting effect.

Systematic multiplicative drift occurs due to drift in the light bulb output over time. In terms of the multiplexing we are interested in drift during the acquisition period. The light drift between complement encodings is assumed very small, but no such assumption

is made about the drift from the first to the last measurement. A reference beam conducts light from the bulb, through a chopper directly to the spectrometer, and a reference beam spectrum measurement is taken immediately after each sample spectrum. The sample spectrum is divided by the corresponding reference beam spectrum before decoding, which removes the multiplicative factor due to light drift. Spectra are taken of a highly reflective white object that is half cardboard and half acrylic. The spectra are corrected by Spectralon white tile spectra. The images are corrected with reference beam correction and without reference beam correction and the difference assessed.

Demonstration that the spectra are useful and sensible is initially examined by imaging simple fabricated objects comprised of different materials with interesting and distinct spectral shapes. Also imaging of a test pattern at different resolutions demonstrates the flexibility of the system. Simple chemometric analysis of the spectra demonstrate that the variation in the spectra is sufficiently low within groups for segmentation of the pixels. We then move on to application of the imaging system to a food application. We perform volumetric mapping of spectra down the length of beef striploins. We test the discrimination of lean from non lean spectral pixels, examine spectral variability down the length of the muscle as well as perform pilot study level analysis of various important chemical features such as acidity, bloom (change in colour from purple to red due to oxygenation of myoglobin over time after initial exposure to air) and tenderometry (measurement of shear force to 'bite' though cooked meat).

## 1.4 Thesis Structure

Chapter 2 comprises the literature review. Multiplexing, relevant hardware configurations and the Hadamard transform are covered, with the compressed sensing method and some main theoretical results included. Chemometrics, that is statistical methods used on spectroscopic data, and other statistical techniques not specific to chemometrics are reviewed. Lastly some results of visible and/or near infrared spectroscopy and chemometrics to analyse beef are covered with some results for pork also included.

Chapter 3 presents the theoretical development for understanding and analysing the hyperspectral imager. The topic of multiplexing is covered in an application independent manner, with the intention of providing a unified framework under which multiplexing matrix error behaviour can be analysed. A model of multiplexing accounting for all reasonably possible error sources is built up. Each error source is considered in turn from which the complete model is then assembled. Theory surrounding the correction of systematic error sources is developed, with the accompanying effect on the random forms of error. The Hadamard H, G and S-matrices are all examined under the framework.

Chapter 3 also covers the extension of the link between strongly regular graphs and

S-matrices to the other Hadamard matrices. Beginning with the graph matrix equivalent to the S-matrix the corresponding G-like and H-like matrices are constructed and analysed. Chapter 3 then moves on to compressed sensing. First we revise pertinent aspects from the literature, then examine the precision behaviour with respect to the number of measurements for a random acquisition matrix. The connection between the Wilcoxon statistic and the area under the receiver operator curve is derived, and the result used to analyse the effect of error in the reference on the Wilcoxon. Numerical simulations support the theoretical result.

Chapter 4 describes the hardware implementation for the hyperspectral imager and interprets the theory in terms of light multiplexing on the system. Methods of testing the hyperspectral imager for Hadamard multiplexing and compressed sensing are given and the results of the testing are presented and discussed.

Chapter 5 is concerned with an application of the imaging system to demonstrate the usability of the spectra produced, namely analysis of beef *m. longissimus dorsi*. Spectral volumetric mapping of the meat was performed with accompanying colour imaging and pH measurement. Bloom curves were also obtained using the colour imaging and tenderometry reading were taken by way of a shear force meter. Classification of the spectra into lean and nonlean groups is performed using the colour imaging to determine visible content as the reference method. The theory of the Wilcoxon statistic with error in the reference is applied to the classification result. Variography is applied to the spectra down the length of the muscle to examine the nature and possible causes of variation in the spectra. Calibration of the spectra against parameters of the start point, end point and rate of bloom is performed. The predictive power of the calibration models is analysed using per-animal crossvalidation. Correlations are explored between the spectra and the pH and tenderometry. Conclusions and thoughts for future work are presented in Chapter 6.

## 1.5 List of Publications

The following publications arose from the research conducted for this thesis.

### 1.5.1 Journal Publications

L. Streeter, G. Burling-Claridge, M. Cree, and R. Künnemeyer. Optical full Hadamard matrix multiplexing and noise effects. *Applied Optics*, **48(11)**, pp. 2078–2085 (2009).

L. Streeter, G. Burling-Claridge, M. Cree, and R. Künnemeyer. Visible/near infrared hyperspectral imaging via spatial illumination source modulation. *Journal of Near Infrared Spectroscopy*, **15(6)**, pp. 395–399 (2007).

### 1.5.2 Conference Papers

L. Streeter, G.R. Burling-Claridge, M.J. Cree and R. Künnemeyer Comparison of Hadamard imaging and compressed sensing for low resolution hyperspectral imaging In: *Image and Vision Computing New Zealand 2008*, pp. 1–6 Christchurch, New Zealand (2008).

L. Streeter, G. Burling-Claridge, M. Cree, and R. Künnemeyer. Reference beam method for source modulated Hadamard multiplexing. *Proc SPIE*, **6812**, pp. 68160J1–10 (2008).

L. Streeter, G. Burling-Claridge, and M. Cree. Image processing of meat images for visible/near infrared spectroscopy reference. In: *Image and Vision Computing New Zealand 2006*, pp. 275–280 Gt Barrier Island, New Zealand (2006).

L. Streeter, P.W. Hansen, R. Burling-Claridge, M.J. Cree, R. Künnemeyer and C.E. Devine Estimation of Visible Fat Content in Beef M. Longissimus Using Near Infrared Spectroscopy In: *52nd International Congress of Meat Science and Technology*, pp. 671 Dublin, Ireland (2006).

L. Streeter, R. Burling-Claridge and M.J. Cree Colour Image Processing and Texture Analysis on Images of Porterhouse Steak Meat In: *Image and Vision Computing New Zealand 2005*, pp. 398–403 Dunedin New Zealand, New Zealand (2005).



# Chapter 2

## Review

This section comprises a review of relevant theory and literature to this thesis. Overview of the fundamentals of multiplexing theory is covered and the traditional and modern multiplexing hardware are reviewed, then the properties of the Hadamard matrix for multiplexing and the newer compressed sensing technique. Near infrared spectroscopy and chemometrics and finally the capture and analysis of spectra and relevant information of the variation in certain beef muscles are reviewed.

### 2.1 Multiplexing

Multiplexing (Harwit and Sloane, 1979) is the technique of grouping objects according to a specific pattern. Specifically, in the present context, multiplexing involves measuring multiple samples simultaneously to improve the signal to noise ratio (SNR) of measurement. Consider a set of  $N$  objects of mass  $x_i$  to be weighed. Now if the weighing device has uncertainty  $\sigma$  then the SNR of each measurement is

$$\frac{x_i}{\sigma}. \tag{2.1}$$

If we weigh half of the objects together at any given time then the SNR is approximately  $N/2$  times greater.

If  $N$  grouped measurements are taken, each time weighing a different set of the objects, then the measurements constitute a linear transformation of the mass values. Let  $M$  be a matrix of ones and zeros that records which objects were included in each measurement, where a one at row  $j$  and column  $i$  indicates that object  $i$  is included in measurement  $j$ . The matrix  $M$  is known as both the multiplexing matrix and the weighing design matrix. Loading the true mass values into the vector  $\mathbf{x}$  the measurement process is

represented by the matrix multiplication

$$\mathbf{y} = M\mathbf{x} + \mathbf{e}, \quad (2.2)$$

where  $\mathbf{y}$  is the set of observed measurements and  $\mathbf{e}$  is random error with variance  $\sigma^2$ . Decoding to recover an estimate of the weights  $\hat{\mathbf{x}}$  is performed by application of the mathematical inverse to the multiplexing matrix

$$\hat{\mathbf{x}} = M^{-1}\mathbf{y} = \mathbf{x} + M^{-1}\mathbf{e}. \quad (2.3)$$

The selection of the matrix  $M$  determines the SNR of the estimates at reconstruction. The error in the reconstruction is  $\hat{\mathbf{x}} - \mathbf{x} = M^{-1}\mathbf{e}$ . If the random errors represented in  $\mathbf{e}$  each have the same variance as each other and  $M$  is suitable then the errors in the reconstruction will also have the same variance as each other, where preferably the error in reconstruction is less than the error in  $\mathbf{e}$ . When the random errors in reconstruction are different to each other some single computable measure of the overall effect of multiplexing on error is required. The effect on average of the squared error of measurements due to multiplexing, called the A-optimality, is such a measure (Harwit and Sloane, 1979). It turns out that the A-optimality of a multiplexing matrix is computed as

$$\epsilon = \frac{\sigma^2}{N} \text{Tr} \left[ (M^T M)^{-1} \right], \quad (2.4)$$

where  $\text{Tr}[\cdot]$  is the matrix trace. A good multiplexing matrix produces a value for  $\epsilon$  that is less than  $\sigma^2$ . The minimum possible value, according to Hotelling's lower bound, is (Harwit and Sloane, 1979)

$$\epsilon = \frac{\sigma^2}{N}. \quad (2.5)$$

## 2.2 Hardware

A spectrometer is a device for measuring the spectral information in light. The term 'light' here is general and pertains to portions of the electromagnetic spectrum that are not visible to the human eye but are close in wavelength to the human visible range, such as ultra-violet and infrared. In the visible and near infrared (Vis/NIR) domain a spectrometer is made from traditional optical elements, such as lenses, gratings and prisms. A spectrometer consists of a slit through which light enters, a dispersion element, such as a grating or prism which separates the light into spectral components and an array of sensors, such as a linear diode array or charged couple device, to measure the light. Figure 2.1 shows a simplified spectrometer. The dispersive element separates the spectral components of the

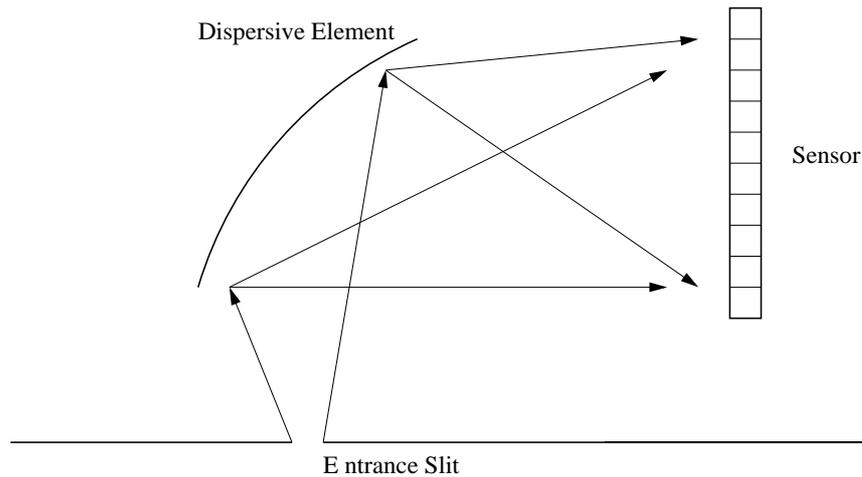


Figure 2.1: Simplified diagram of a spectrometer. The arrows indicate the light path. Light entering through the entrance slit is dispersed and reflected onto the sensor.

light so that each sensor element is exposed to unique bandwidth (typically labelled as a wavelength) of light. A light source typically accompanies the spectrometer. Tungsten halogen bulbs have broad spectral blackbody output with good temporal stability so are suitable for Vis/NIR spectroscopy. Light from the source reaches the entrance slit by two routes. The first route is light reflected off the sample. The second route is via a reference beam that conducts as much light directly from the light source as possible without saturating the sensor.

The concept of light multiplexing to boost signal to noise for use in the context of spectrometry is attributed to Golay (1949). Harwit and Sloane (1979) state that “Golay’s work more than any other laid the foundation for multiplexed spectroscopy”. The early multiplexed spectrometers had multiple slits, known as multislit configurations. Golay first proposed a spectrometer with two masks that form a virtual multislit system. One mask was a rotating disc and the other a rectangular mask. The rotating disc had encoding patterns in the form of holes and occlusions that encoded spectroscopic information. The spectral signal was later reconstructed via computer processing.

Traditional optical light multiplexing systems implemented the S-matrix encoding (Harwit and Sloane, 1979), Golay patterns (Golay, 1949) and random sequences of maximal length (Wilhelmi and Gompf, 1970). Other traditional examples are surveyed by Harwit and Sloane (1979). Encoding was performed by occlusion and passage of light. Figure 2.2 shows a diagram illustrating the principle of multislit spectrometry. Light enters through the entrance slit and is separated into spectral components. A mechanical encoding multislit element selects which spectral components reach the sensor. The combination of spectral components that reach the sensor results in more light per single measurement than when each spectral component is measured individually. The encoding element is changed between successive measurements, resulting in a linear transformation

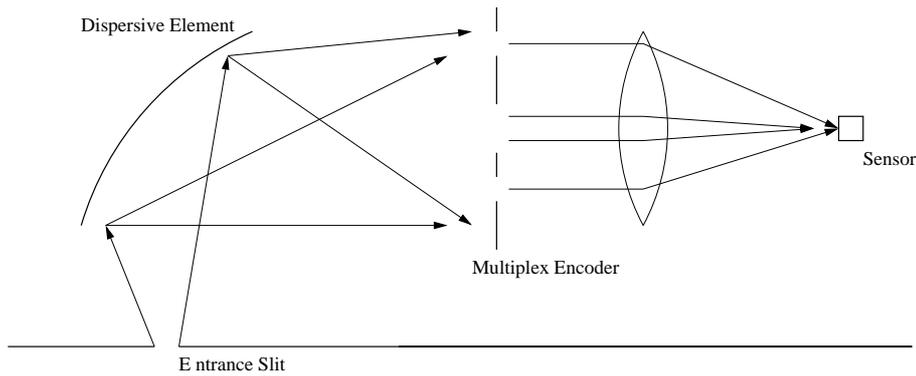


Figure 2.2: Diagram of a multiplexing spectrometer. The arrows indicate the light path. Spectral wavelength components separated by the dispersion element are selectively passed and occluded by the encoder and measured by the sensor. Successive measurements utilise different sets of wavelengths, encoding the spectrum.

of the spectrum. Mechanical systems to perform such light modulation include multislit spectrometers, one and two dimensional rectangular grating arrays, rotating discs and combinations thereof. Multiplex encoding using holes and occlusions has the obvious disadvantage that a large fraction of the light from the sample is not used. Sloane *et al.* (1969) pointed out that the fraction of light normally lost could be recovered via reflection in place of occlusion. Light sensors then measure both the transmitted and reflected light. Subtraction of the two signals is implemented electronically using a subtractor circuit or digitally after data capture. This concept was later extended by Davis (1995) to two dimensions via two one-dimensional encodings to optically implement the Kronecker product.

The advent of the Texas Instruments Digital Micromirror Array (DMA) (Hornbeck, 1989) has facilitated the revision and revitalisation of optical multiplexing techniques. The DMA is an array of small (typically  $14\mu\text{m} \times 14\mu\text{m}$ ) mirrors that can be individually addressed to deflect to a nominally on or off state. Example DMA sizes are 0.55" or 0.7" diagonal with respectively  $600 \times 800$  or  $1024 \times 768$  mirrors. DMAs are sufficiently fast (the slower off-the-shelf evaluation kit boasts 100 fps via USB interface) and flexible. Also the broad spectral characteristics of mirrors make the DMA a natural choice for Vis/NIR spectroscopic applications. The DMA is reliable and due to their small size the mirrors do not resonate with the mechanical vibration frequencies of a typical working environment.

Figure 2.3 shows a diagram of a generic DMA based imaging spectrometer system similar to that proposed by DeVerse *et al.* (2000). The light from the entrance slit is dispersed in one axis and imaged in the other onto the DMA. The DMA then modulates spatial and/or spectral information and the modulated light is sensed by the CCD. The two dimensional CCD is interchangeable with a one dimensional linear array or a

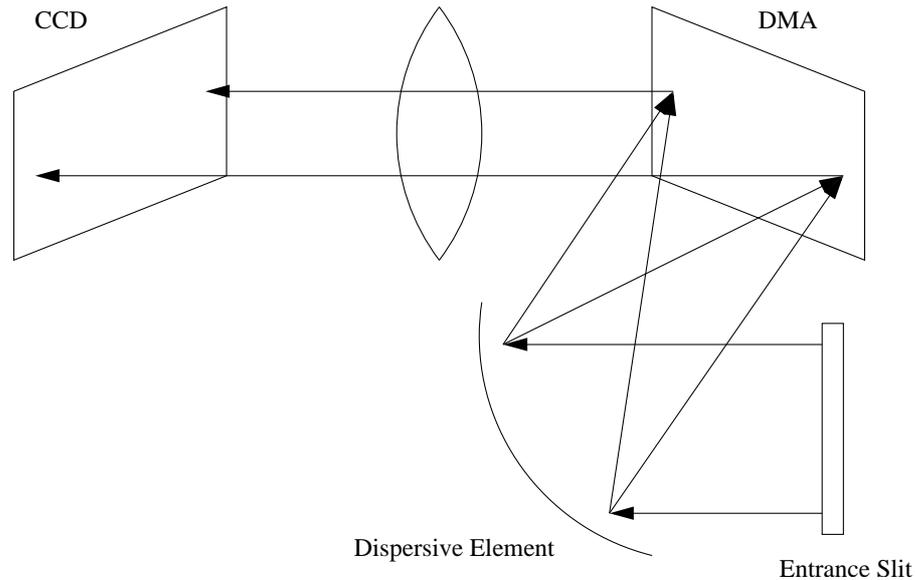


Figure 2.3: DMA and CCD based imaging spectrometer. The DMA modulates the spectral and/or spatial information which is then recorded by the CCD.

single point diode for the specific mode of multiplexing and imaging desired. The spectral dispersion can be implemented before the sample (de-dispersion) or after the sample (dispersion). The use of the DMA in multiplexing was implemented in dispersive NIR spectrometry (Fateley *et al.*, 2000; DeVerse *et al.*, 2000), de-dispersive Vis/NIR spectroscopy (Fateley *et al.*, 2002), Raman imaging (DeVerse *et al.*, 1999, 2000) and visible spectrometry (Spudich *et al.*, 2003) using a combination of DMA based modulation of spectral and/or spatial information coupled with two dimensional (CCD or similar) sensor technology. A de-dispersive system using a single diode sensor was described by Fateley *et al.* (2002) in which the DMA was directly illuminated by the broadband source, the light spectrally dispersed and imaged onto the sample so that the spectral and one spatial dimension is captured. The sample was then translated to build up the second spatial dimension. The hyperspectral imaging was applied to the detection of cancerous tissue (Maggioni *et al.*, 2006).

A DMA based microscope (Hanley *et al.*, 1998) utilised the Hadamard transform to implement optical sectioning, a method for improving image resolution similar to confocal microscopy but with greater light throughput (Verveer *et al.*, 1998). A number of pinholes are used in patterns so that light from the plane of interest is encoded. Out of focus light from out of the plane of interest is equally present in each capture. Decoding, which involves a series of additions and subtractions, then reconstructs the detail of the sectioned plane and removes the light from out of the plane. Spatial and spectral resolution is obtained with a CCD based imaging spectrograph that examines one spatial and the spectral dimension. Mechanical pushbroom scanning facilitates acquisition of the

second dimension. A similar optical configuration (Hanley *et al.*, 2000) utilised a transmissive LCD in place of the DMA. The LCD microscope implemented multiplexing with pushbroom scanning. A further adaptation (Hanley and Jovin, 2001) implemented optical sectioning with multiplexing using a CCD based spectrometer.

## 2.3 Hadamard Transform

The Hadamard matrices are ideal for multiplexing under certain conditions. Here we review the construction and properties of Hadamard matrices. Consider the problem of finding a square matrix with  $N$  rows and columns, where each entry is real and has absolute value less than or equal to one, and the matrix has the largest possible determinant. The Hadamard H-matrix (Harwit and Sloane, 1979) is the solution, the construction of which was first described by Sylvester (1867). The Sylvester construction first defines the solution for the  $1 \times 1$  matrix,  $H_1$ , and the  $2 \times 2$  matrix,  $H_2$ , as

$$\begin{aligned} H_1 &= [1], \\ H_2 &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \end{aligned} \quad (2.6)$$

Given two H-matrices  $H_N$  and  $H_M$ , a new H-matrix  $H_{N+M}$  is constructed by taking the Kronecker product, that is, each element in  $H_M$  is replaced by that element times  $H_N$ . For example the matrix  $H_{2+2} = H_4$  is

$$\begin{aligned} H_4 &= \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \end{aligned} \quad (2.7)$$

The examples above all have ones on the leading row and column. Such H-matrices are said to be normalised. Rearranging the rows or columns can produce equally valid unnormalised H-matrices, for example

$$\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}. \quad (2.8)$$

The Hadamard transform is the linear transform according to the H-matrix. For a

vector  $\mathbf{x}$  the transformation is the matrix product

$$H\mathbf{x}. \quad (2.9)$$

The rows (and columns) of the H-matrix are pairwise orthogonal. The vector product of any two rows (or columns) in  $H_N$  is zero, but if the same row is chosen twice then the product is  $N$ . Written succinctly, where  $\cdot^T$  is the matrix transpose operation and  $I_N$  is the matrix identity,

$$H_N^T H_N = H_N H_N^T = N I_N. \quad (2.10)$$

From equation 2.10 the inverse of the H-matrix is  $H^T/N$ , thus the inverse transformation gives

$$\mathbf{x} = \frac{1}{N} H^T H \mathbf{x}. \quad (2.11)$$

For the normalised H-matrices above  $H^T = H$  therefore  $H^{-1} = H/N$ , but this is not true for the non-normalised example. The self transpose property is useful so heretofore we only consider normalised H-matrices.

The H-matrix has optimal matrix condition and thus is highly insensitive to random perturbations such as noise. Adding error,  $\mathbf{e}$ , to the Hadamard transform and taking the inverse transformation gives an estimate  $\hat{\mathbf{x}}$

$$\begin{aligned} \hat{\mathbf{x}} &= \frac{1}{N} H (H \mathbf{x} + \mathbf{e}), \\ &= \mathbf{x} + \frac{1}{N} H \mathbf{e}. \end{aligned} \quad (2.12)$$

The A-optimality of the Hadamard matrix is (Harwit and Sloane, 1979)

$$\begin{aligned} \epsilon &= \frac{\sigma^2}{N} \text{Tr} \left[ (H^T H)^{-1} \right], \\ &= \frac{\sigma^2}{N} \text{Tr} \left[ (N I)^{-1} \right], \\ &= \frac{\sigma^2}{N}. \end{aligned} \quad (2.13)$$

This is the minimum achievable according to Hotelling's bound on the error. The reduction of additive noise is called the Hadamard multiplex advantage. It is closely related to the Fellgett advantage, which pertains to multiplexing according to the Fourier transform.

There are negative entries in the H-matrix which are problematic (but not impossible) for implementation in optical systems. The S-matrix (simplex matrix) is an alternative which has no negative entries. The S-matrix can be derived from the H-matrix (there

exist other construction methods). First the G-matrix,  $G_{N-1}$ , is defined as

$$H_N = \begin{bmatrix} 1 & \mathbf{0}_{N-1}^T \\ \mathbf{0}_{N-1} & G_{N-1} \end{bmatrix}, \quad (2.14)$$

where  $\mathbf{0}_{N-1}$  is a column vector of  $N-1$  ones. From equation 2.6 it is apparent that  $G_0 = []$ , an empty matrix. The S-matrix is obtained by setting the ones in G-matrix to zeros and the negative ones to positive ones. Alternatively we can simply set all the negative ones to zeros (although in using this alternative method we must be aware that we have the complement of the S-matrix). About half of each row of  $S$  is ones and the other half zeros, so the S-matrix encoding employs an approximately 50% duty measurement system while maintaining maximal independence between rows (and columns). The inverse of the S-matrix is

$$S_{N-1}^{-1} = \frac{2}{N} G_{N-1}. \quad (2.15)$$

The A-optimality of S-matrix encoding is

$$\begin{aligned} \epsilon &= \frac{\sigma^2}{N} \text{Tr} \left[ (S^T S)^{-1} \right], \\ &= \frac{\sigma^2}{N} \text{Tr} \left[ \frac{4}{N^2} G^2 \right], \\ &\approx \frac{4\sigma^2}{N}. \end{aligned} \quad (2.16)$$

Thus the S-matrix transform reduces the average MSE by a factor of  $4/N$ . The S-matrix is conjectured to be optimal in terms of insensitivity to random additive noise for all matrices of ones and zeros, however the S-matrix does not carry quite the same promise of signal improvement that the H-matrix does.

In optical systems Poisson photon noise effects are known to be detrimental to the advantage of Hadamard encoding, particularly in S-matrix encoding (Nitzsche and Riesen-berg, 2003; Damaschini, 1993). Such noise effects are treated as multiplicative, i.e the noise variance at acquisition is directly proportional to the light level. The S-matrix permits about 50% of light to reach the sensor, so the noise variance at data acquisition is then proportional to

$$\approx N \langle x \rangle / 2, \quad (2.17)$$

where  $\langle x \rangle$  is the mean of the entries in  $\mathbf{x}$ . Inverting the S-matrix transform then gives a noise variance proportional to

$$\approx 2 \langle x \rangle. \quad (2.18)$$

In contrast if no transformation is used then the noise for each entry in  $\mathbf{x}$  is proportional to that entry. On average the noise is proportional to  $\langle x \rangle$ , half that of S-matrix multiplexing,

so S-matrix multiplexing is disadvantageous for signal proportional noise.

Hassler *et al.* (2005) studied the effect of H-matrix multiplexing in the context of multiplexed time resolved fluorescence imaging. In such imaging a sample is exposed to a series of excitation light flashes and the time for the the first photon from the fluorescence to reach the sensor is recorded for each excitation. The time for the light to reach the sensor follows Poisson statistics. A histogram of the times recorded is obtained to which a Poisson distribution is fit. Imaging was performed by Hadamard multiplexing of the light from the sample to the sensor. Hassler *et al.* (2005) presented theoretical analysis in which each ‘pixel’ region on the object was treated as a separate emitter, and the light from the pixels was then selectively passed or obstructed and combined at the image sensor. They derived theoretically that on average the photon noise level in the pixel reconstruction was not changed by the multiplexing. They point out that the averaging of the Poisson noise level by H-matrix encoding will increase the noise in dark pixels which may be unhelpful in some applications.

Alternative derivations in the context of mass spectrometry have indicated a SNR boost under S-matrix encoding with photon noise (Wilhelmi and Gompf, 1970; Fernández *et al.*, 2001). Mass spectra typically are sparse, meaning that there are few peaks and most of the measurements are indistinguishable from zero. The signal to noise ratio ‘spectrum’ associated with a mass spectrum is indistinguishable from zero except at the locations of the peaks. It is not appropriate to consider the boost in the signal to noise ratio due to multiplexing at the zero locations because it involves the ratio of two numbers that are indistinguishable from zero. An alternative assessment involves examination of the SNR at the peaks in the spectra only.

Alternative encoding strategies have been proposed (Wilhelmi and Gompf, 1970; Wuttig, 2005; Ratner *et al.*, 2007) to provide optimal noise reduction under photon noise or to trade off between reduction of photon noise and instrument noise. In general these photon noise reduction transformations employ less than 50% duty while maintaining maximum possible independence between rows in the encoding matrix. In particular Ratner *et al.* (2007) utilise a result from graph theory, namely the strongly regular graph (SRG). Consider a graph of  $N$  vertices (points), where each vertex is connected to  $k$  other vertices. The connections are called edges and two points that are connected are said to be adjacent and are called neighbours. Now assume that each adjacent pair of vertices has  $\alpha$  common neighbours. Furthermore assume that each nonadjacent pair has  $\beta$  common neighbours. Such a graph is called strongly regular and is denoted  $\text{srg}(N, k, \alpha, \beta)$ . Any graph is represented by its adjacency matrix where each row and column corresponds to each vertex in the graph. A one at position  $(i, j)$  indicates that vertices  $i$  and  $j$  are neighbours whereas a zero indicates that they are not. It turns out that when  $\beta = \alpha$  the eigenvalue structure of the adjacency matrix matches that of the S-matrix of the same

size. The eigenvalue structure is linked to the noise behaviour due to multiplexing. Furthermore a SRG does not require each vertex to be adjacent to half the other vertices, so the matrix in question does not necessarily require half the entries in each row to be ones. Thus when multiplexing via an appropriate adjacency matrix the signal and photon noise variance is reduced. When the signal is reduced then the signal boost over the additive noise is also reduced so there is a trade off between the advantage over additive and multiplicative noise. Given known additive and multiplicative noise levels Ratner *et al.* (2007) show how  $k$  can be selected to minimise the total noise. Specifically if the additive noise variance is  $\sigma^2$  and the photon noise per measurement ( $x_i$  in  $\mathbf{x}$ ) is  $\rho$ , then multiplexing by an appropriate  $N \times N$  adjacency matrix results in total noise MSE

$$\frac{\sigma^2 + \rho k}{N} \left( \frac{1}{k^2} + \frac{(N-1)^2}{Nk - k^2} \right), \quad (2.19)$$

at decoding. Optimising the multiplexing is a matter of finding  $k$  so that equation 2.19 is minimised. For a given  $N$  and  $k$ , there may not be a SRG with  $\beta = \alpha$ . However there are many known strongly regular graphs from which one can select the one with the closest parameters. Furthermore SRGs exist for values of  $N$  for which no known Hadamard matrix exists.

## 2.4 Compressed Sensing

Compressed sensing, or compressive sensing (CS), is the technique of reconstructing a discrete signal from a reduced number of measurements when the signal is compressible (Candes *et al.*, 2006; Donoho, 2006). Random multiplexing patterns,  $\Phi$ , are used to acquire data

$$\mathbf{y} = \Phi \mathbf{x}, \quad (2.20)$$

where  $\Phi$  has fewer rows than columns. The signal  $\mathbf{x}$  is assumed sparse in some transformation  $\Psi$ . In other words if  $\mathbf{x}$  is represented by the reconstruction

$$\mathbf{x} = \Psi \boldsymbol{\theta}, \quad (2.21)$$

then the vector  $\boldsymbol{\theta}$  is a set of coefficients where a few are large but most are zero or near zero. The system to be solved is then

$$\mathbf{y} = \Phi \Psi \boldsymbol{\theta}, \quad (2.22)$$

from which the estimate of  $\mathbf{x}$  is computed.

Consider an acquisition basis,  $\Phi$ . Let  $\Phi_T$  be any subset of at most  $s$  columns in  $\Phi$ , then

the restricted isometry constraint,  $\delta_s \in (0, 1)$ , is the smallest number such that (Candès and Tao, 2005)

$$(1 - \delta_s)\|\mathbf{p}\|_2^2 \leq \|\Phi_T \mathbf{p}\|_2^2 \leq (1 + \delta_s)\|\mathbf{p}\|_2^2, \quad (2.23)$$

where  $\|\cdot\|_2$  is the L2 norm. A matrix  $\Phi$  that satisfies equation 2.23 is said to satisfy the restricted isometry property (RIP) (Baraniuk *et al.*, 2008). An equivalent statement of the RIP is to replace  $\Phi_T$  with  $\Phi$  and consider  $s$  sparse vectors  $\mathbf{p}_s$  which have at most  $s$  nonzero entries (Candès, 2008). It turns out that random matrices satisfy the RIP and form suitable CS acquisition systems. The RIP states that a subset of the columns in  $\Phi$  can alter the L2 norm of any vector by at most a small fraction. CS seeks a representation of  $\mathbf{p}$  as a combination of basis vectors. The RIP ensures that the relative magnitude of the contributions of these basis vectors is minimally altered at encoding.

Reconstruction of  $\theta$  is done via L1 norm minimisation. The estimate is found by solving the linear programming problem

$$\min\|\theta\|_1 \text{ subject to } \mathbf{a} = \Phi\Psi\theta, \quad (2.24)$$

where  $\|\cdot\|_1$  is the L1 norm (the sum of the absolute values). The linear programming problem is solved by the basis pursuit algorithm<sup>1</sup> (Chen *et al.*, 1998), among others.

Basis pursuit uses a primal-dual log-barrier linear programming algorithm (Chen *et al.*, 1998). The approach involves solving the perturbed linear program proposed by Gill *et al.* (1991):

$$\min \mathbf{c}^T \mathbf{x} + \frac{1}{2}\|\gamma \mathbf{x}\|^2 + \frac{1}{2}\|\mathbf{q}\|^2 \text{ subject to } A\mathbf{x} + \delta \mathbf{q} = \mathbf{b}, \mathbf{x} \geq 0, \quad (2.25)$$

where  $\mathbf{x}$  is the unknown of interest,  $\mathbf{b}$  is the measured data and  $\gamma$  and  $\delta$  are small regularisation parameters (Chen *et al.* (1998) suggest  $10^{-4}$ ). The procedure given by Chen *et al.* (1998) is as follows.

1. Set the feasibility tolerance FT, the duality gap tolerance DGT,  $\gamma$  and  $\delta$ .
2. Initialise the prime variable,  $\mathbf{x} > 0$ , the dual variables,  $\mathbf{y} = 0$  and  $\mathbf{z} > 0$ , and  $\mu > 0$ .
3. Loop

---

<sup>1</sup>SparseLab Matlab package Version:100, implementation downloaded from <http://sparselab.stanford.edu/>, August 2007, Copyright (c) 2006 Victoria Stodden and David Donoho.

(a) Compute the residuals and the diagonal matrix  $D$  by

$$\begin{aligned}\mathbf{t} &= \mathbf{c} + \gamma^2 \mathbf{z} - \mathbf{z} - A^T \mathbf{y} \\ \mathbf{r} &= \mathbf{b} - A\mathbf{x} - \delta^2 \mathbf{y} \\ \mathbf{v} &= \mu \mathbf{o} - Z\mathbf{x} \\ D &= (X^{-1}Z + \gamma^2 I)^{-1}\end{aligned}$$

where  $X$  and  $Z$  are diagonal matrices formed from  $\mathbf{x}$  and  $\mathbf{z}$ .

(b) Solve for  $\Delta \mathbf{y}$

$$(ADA^T + \delta^2 I)\Delta \mathbf{y} = \mathbf{r} + AD(\mathbf{t} - X^{-1}\mathbf{v})$$

and compute

$$\Delta \mathbf{x} = D(A^T \Delta \mathbf{y} + X^{-1}\mathbf{v} - \mathbf{t}) \text{ and } \Delta \mathbf{z} = X^{-1}(\mathbf{v} - Z\Delta \mathbf{x})$$

(c) Calculate the primal and dual step sizes  $\rho_p$  and  $\rho_d$ , and update

$$\begin{aligned}\rho_p &= 0.99 \max\{\rho|\mathbf{x} + \rho\Delta \mathbf{x}\} \\ \rho_d &= 0.99 \max\{\rho|\mathbf{z} + \rho\Delta \mathbf{z}\} \\ \mathbf{x} &= \mathbf{x} + \rho_p \Delta \mathbf{x}, \mathbf{y} = \mathbf{y} + \rho_d \Delta \mathbf{y}, \mathbf{z} = \mathbf{z} + \rho_d \Delta \mathbf{z} \\ \mu &= (1 - \min(\rho_p, \rho_d, 0.99))\mu\end{aligned}$$

4. Terminate if the following are all satisfied

- (a)  $\frac{\|\mathbf{r}\|_2}{1+\|\mathbf{x}\|_2} < \text{FT}$
- (b)  $\frac{\|\mathbf{t}\|_2}{1+\|\mathbf{y}\|_2} < \text{FT}$
- (c)  $\frac{\mathbf{z}^T \mathbf{x}}{1+\|\mathbf{z}\|_2\|\mathbf{x}\|_2} < \text{DGT}$

For basis pursuit set  $\mathbf{c} = \mathbf{o}$ , a vector of ones, to implement the L1 norm. To satisfy  $\mathbf{x} \geq 0$  solve the system matrix  $[A, -A]$  for  $\mathbf{x}$ . The result is the length  $2N$  vector comprising of the concatenated positive and negative parts of the desired solution  $\mathbf{x}_+$  and  $\mathbf{x}_-$  from which  $\hat{\mathbf{p}} = \mathbf{x}_+ - \mathbf{x}_-$  is computed.

The error of CS in the noiseless case is bounded (Donoho, 2006), for L1 norm minimisation the upper bound is

$$C\|\theta\|_1 \left(\frac{\log N}{n}\right)^{1/2} \quad (2.26)$$

where  $n$  (the number of rows in  $\Phi$ ) gets large,  $C$  is a constant that depends on at most  $\log N / \log n$  and  $\|\cdot\|_1$  is the L1 norm. Furthermore CS is stable against additive random noise (Candes *et al.*, 2006; Haupt and Nowak, 2006) and the combination of additive and

multiplicative random noise (Haupt *et al.*, 2006).

An implementation of CS for imaging is the single pixel camera (Takhar *et al.*, 2006; Wakin *et al.*, 2006a). The single pixel camera is a passive imaging system performing CS on light from the object. A DMA spatially modulates pixel information using randomly assigned deflection of light to or away from a light sensitive diode. Reconstruction of the image is performed using L1 norm minimisation with the fast wavelet transform as the compression basis. Video imaging has been implemented on the single pixel camera (Wakin *et al.*, 2006b). Time domain CS assisted in improving the capture rate over CS in the spatial domain only. CS has the obvious advantage that fewer measurements provides faster acquisition than traditional techniques like Hadamard imaging, however it is not clear how CS compares to Hadamard imaging in terms of signal quality. The relative performance of CS and Hadamard imaging are compared in section 4.3.3.

A single-shot CS hyperspectral imager, described by Gehm *et al.* (2007), used symmetric optics and a CCD imager. The optics has two identical arms where each arm is a dispersive spectrometer that uses an equilateral prism. The second arm effectively cancels out the dispersive action of the first arm. Between the two arms is a coding aperture based on an S-matrix or order 192. The image of an object is spectrally dispersed by the first optical arm. The dispersed image is then encoded and undispersed on the second arm. The result is captured by the CCD imager. The hyperspectral image is reconstructed via a probabilistic multiresolution framework. Spectral imaging was performed in the range 520–590nm, with filtering to prevent detection of light of other wavelengths. Wagadarikar *et al.* (2007) present an alternative single shot architecture which uses one dispersive arm and the coding aperture. Again the CS paradigm was invoked, where fewer measurements than reconstructed data values were measured. Spectral imaging was achieved in the range 540 – 640nm. The object is imaged onto the coded aperture. Relay lenses transmit light from the coded aperture, through the dispersive element and reimage onto the CCD. The single-shot paradigm, facilitated by use of a CCD sensor, requires one capture rather than the multiple captures required for traditional multiplexing, however the spectra obtained are over a narrow band whereas the use of a traditional spectrometer permits much broader spectral analysis.

## 2.5 Near Infrared Spectroscopy and Chemometrics

Chemometrics is the application of specialised statistical techniques to spectroscopic data. Spectroscopic data include, but are not limited to, photon spectroscopy of which the visible and near infrared spectrum is of interest to this thesis (Osborne *et al.*, 1993). The objective of chemometric analysis is often to calibrate the output of a spectrometer with some reference substance concentration or effect, or to detect the presence of a substance

or effect by classification.

The simplest chemometric methods include construction of a linear regression model against a selected subset of wavelength responses, or against all wavelengths measured. In spectroscopy there is often a large number of interdependent measured observables, so over fitting often occurs in linear regression. Wavelength selection is one method to overcome over fitting. Factor analysis methods, otherwise viewed as subspace analysis methods, are full spectrum techniques that “aim to reduce the quantity of spectral data, and thus avoid overfitting problems, without discarding any useful information” (Osborne *et al.*, 1993). The most prominent factor analysis methods are principle component regression (PCR) (Osborne *et al.*, 1993) and partial least squares (PLS) (Wold, 1975; de Jong, 1993). PCR exploits eigenvector analysis of the variance-covariance structure of the spectral data only. PLS computes factors based on the covariance between the spectral data and the reference data used for calibration. Nonlinear techniques include neural networks (Osborne *et al.*, 1993) and support vector regression (Smola and Schölkopf, 2004). Classification in chemometrics can be performed directly using PLS (Barker and Rayens, 2003) or alternatively via extended canonical variates analysis (Nørgaard *et al.*, 2006), an application of PLS to find the solution to Fisher linear discriminant analysis.

### 2.5.1 PCR and PLS

Let  $X$  be the  $n \times m$  matrix of spectral data with each row a spectrum and each column a narrow waveband and  $\mathbf{y}$  the  $n \times 1$  vector containing the reference variable. Both PCR and PLS seek a  $m \times k$  matrix of basis vectors  $W$ , where  $k$  is small (typically less than 10) so that the fit

$$\mathbf{y} = XW\mathbf{c}, \quad (2.27)$$

where  $\mathbf{c}$  is a  $k \times 1$  vector, is robust and accurate. Both PCR and PLS are factor analysis methods for modelling a relationship between multivariate data and some reference variable when the multivariate data is highly collinear. Collinearity means that the columns in  $X$  are pairwise highly correlated which drastically impedes the performance of traditional least squares regression methods. In the context of spectroscopy, collinearity means that the response of the measured wavelengths give similar information. For PCR  $W$  is a subset of the eigenvectors of  $X^T X$  chosen to meet some criterion, like capturing most of the variation in  $X$ .

Partial least squares (PLS) (de Jong, 1993) computes  $W$  by examination of the covariance structure between the data,  $X$ , with the reference variable,  $y$ . PLS algorithms exist for multivariate reference data but here we restrict ourselves to the single reference variable case. Consider a reference variable vector  $\mathbf{y}$  of size  $n \times 1$  and a data variable

matrix  $X$  of size  $n \times m$ . Assume a linear model of the form,

$$\mathbf{y} = X\mathbf{b}. \quad (2.28)$$

PLS factorises  $X$  and  $\mathbf{y}$  according to the  $n \times k$ ,  $k < m$ , score matrix  $T$

$$\begin{aligned} X &= TP + E, \\ \mathbf{y} &= T\mathbf{q} + \mathbf{f}, \\ T &= XW. \end{aligned} \quad (2.29)$$

where  $P$  and  $\mathbf{q}$  are loading variables,  $W$  is a matrix of  $k$  weighting vectors of length  $m$  and  $E$  and  $\mathbf{f}$  are unexplained information. Some PLS algorithms make the columns of  $T$  orthogonal whereas others make the columns of  $W$  orthogonal. The objective of PLS is to find  $W$ , hence  $T$ , with small  $k$  so that a stable relationship between  $X$  and  $\mathbf{y}$  is found with minimal risk of overfitting

$$\mathbf{y} = T\mathbf{c}. \quad (2.30)$$

From the last line of equation 2.29

$$\mathbf{y} = T\mathbf{c} = XW\mathbf{c}, \quad (2.31)$$

$$\Rightarrow \mathbf{b} = W\mathbf{c}. \quad (2.32)$$

The PLS algorithm proceeds as follows. The data are mean centred per column,  $X_0 = X - \bar{\mathbf{x}}$  and  $\mathbf{y}_0 = \mathbf{y} - \bar{y}$ . If a very large number of measurements are represented in  $X$  and  $\mathbf{y}$  ( $n$  is large) then for computational efficiency one may replace  $X_0$  and  $\mathbf{y}_0$  with the covariances  $X_0^T X_0/n$  and  $X_0^T \mathbf{y}_0/n$ . At each  $i$ th iteration a new weight vector,  $\mathbf{w}_i$ , is found, from which a new score vector,  $\mathbf{t}_i$ , and loadings,  $\mathbf{p}_i$  and  $q_i$  are computed, the  $X$  and  $\mathbf{y}$  data are deflated to remove the factor corresponding to  $\mathbf{w}_i$  and the deflated  $X$  and  $\mathbf{y}$  then become the subject of the next iteration. The following is iterated  $k$  times.

1. Compute  $\mathbf{w}_i = X_{i-1}^T \mathbf{y}_{i-1} / \|X_{i-1}^T \mathbf{y}_{i-1}\|$  and load  $\mathbf{w}_i$  into the  $i$ th column of  $W$ .
2. Compute  $\mathbf{t}_i = X_{i-1} \mathbf{w}_i$  and load  $\mathbf{t}_i$  into the  $i$ th column of  $T$ .
3. Compute  $\mathbf{p}_i = \mathbf{t}_i^T X_{i-1} / \|\mathbf{t}_i\|^2$  and  $q_i = \mathbf{t}_i^T \mathbf{y}_{i-1} / \|\mathbf{t}_i\|^2$  and respectively load into the  $i$ th row of  $P$  and  $\mathbf{q}$ .
4. Compute the residuals  $X_i = X_{i-1} - \mathbf{t}_i \mathbf{p}_i$  and  $\mathbf{y}_i = \mathbf{y}_{i-1} - \mathbf{t}_i q_i$ . The residuals become the new  $X$  and  $\mathbf{y}$  for the next iteration.

In the above the scores,  $\mathbf{t}_i$ , carry the variance represented in each factor where  $\mathbf{w}_i$ ,  $\mathbf{p}_i$  and  $q_i$  are normalised. One can use  $P$  to predict  $\mathbf{q}$  and subsequently reconstruct  $\mathbf{y}$ . Alternately

we can recompute  $T = XW$  without the mean centring and solve equation 2.30 using least squares regression, namely

$$\mathbf{c} = (T^T T)^{-1} T^T \mathbf{y}. \quad (2.33)$$

The vector  $\mathbf{c}$  is then used with the computed  $W$  to predict  $\mathbf{y}$ . Equation 2.33 is the method we use for the data analysis in this thesis.

## 2.5.2 Spectral Preprocessing

Spectral preprocessing can make chemometric analysis more robust. Unwanted light or sample contamination can influence spectral features negatively, confounding chemometric analysis. First or second order numerical derivatives (Osborne *et al.*, 1993) were traditionally used to remove baseline offset and slope and make spectral peaks appear sharper. The standard normal variate transform (SNV) (Barnes *et al.*, 1989) and multiplicative signal correction (MSC) (Martens *et al.*, 1983) respectively normalise per spectrum and per data set for gross spectral effects. Let the  $i$ th spectrum in data set  $X$  be  $\mathbf{x}_i$  with mean  $\bar{x}_i$  and standard deviation  $\sigma_i$  of the values in the vector  $\mathbf{x}_i$ . The standard normal variate transform sets the mean of each spectrum to zero and the standard deviation to one by

$$\mathbf{x}_{i,\text{SNV}} = \frac{\mathbf{x}_i - \bar{x}_i}{\sigma}. \quad (2.34)$$

After performing the correction a detrend operation is frequently used, in which a straight line is fit to each corrected spectrum and subsequently subtracted. Now let the mean spectrum be  $\bar{\mathbf{x}}$ , the MSC models the spectra as

$$\mathbf{x}_i = a_i + b_i \bar{\mathbf{x}}. \quad (2.35)$$

The parameters  $a_i$  and  $b_i$  are estimated by regression for each spectrum over the wavelengths. The correction is then

$$\mathbf{x}_{i,\text{MSC}} = \frac{\mathbf{x}_i - a_i}{b_i}. \quad (2.36)$$

A mathematical link between SNV and MSC has been established (Dhanoa *et al.*, 1994), where by making the appropriate substitutions one can transform from SNV corrected spectra to MSC.

Extended MSC (EMSC) (Martens *et al.*, 2003) includes polynomial factors in the spectral ‘x-axis’ (wavelength) and reference spectral factors as determined by the user, who can choose what to remove from the spectra and what to leave in. Let  $\mathbf{s}_j$  be representative spectra chosen by the user,  $\lambda$  the wavelength index and  $k_j$  a set of coefficients. The

EMSC model is

$$\mathbf{x} = a_i + b_i \bar{\mathbf{x}} + k_{i,1} \lambda + k_{i,2} \lambda^2 + \sum_j k_{i,j} \mathbf{s}_j. \quad (2.37)$$

The  $\lambda$  and  $\lambda^2$  terms represent straight line and parabola terms. The coefficients  $a_i$ ,  $b_i$  and  $k_{i,j}$  are estimated for each spectrum by least squares.

### 2.5.3 Extended Canonical Variates Analysis

Classification is the algorithmic separation of data into two or more groups, or classes. Extended canonical variates analysis (ECVA) (Nørgaard *et al.*, 2006) is an application of PLS to solve the Fisher linear discriminant analysis (LDA) classification method when the data,  $X$ , are highly collinear. LDA seeks a vector  $\mathbf{b}$  which separates the data according to two or more classes listed in the reference variable  $\mathbf{y}$ . Let  $X_i$  be the submatrix of  $X$  with all instances of class  $i$ . Let  $\mathbf{m}$  be the vector difference between the means of class pairs and  $S_i$  be the scatter (variance-covariance) of  $X_i$ . For the two class problem, Fisher LDA poses the scalar objective function to be maximised as

$$J(\mathbf{b}) = \frac{\mathbf{b}^T S_b \mathbf{b}}{\mathbf{b}^T S_w \mathbf{b}}, \quad (2.38)$$

where  $S_b$  is the between class scatter matrix given by

$$S_b = \mathbf{m}^T \mathbf{m}, \quad (2.39)$$

and  $S_w$  is the within class scatter

$$S_w = \sum_i S_i. \quad (2.40)$$

The solution to equation 2.38 is the solution to the generalised eigenvalue problem with the largest eigenvalue, that is

$$S_w \mathbf{b} = \lambda S_b \mathbf{b}. \quad (2.41)$$

However, using equation 2.39, the right hand side of equation 2.41 is

$$\lambda S_b \mathbf{b} = \lambda \mathbf{m}^T \mathbf{m} \mathbf{b}. \quad (2.42)$$

The product  $\lambda \mathbf{m} \mathbf{b}$  is a scalar. Since LDA is concerned with finding the optimal vector direction scalar terms are inconsequential so  $J(\mathbf{b})$  is maximised by the solution to the linear system

$$S_w \mathbf{b} = \mathbf{m}^T. \quad (2.43)$$

When  $X$  is highly collinear then  $S_w$  is poorly determined. ECVA uses PLS to find a robust solution to equation 2.43. Nørgaard *et al.* (2006) go further and expand the above for three or more classes, here we need only consider the two class problem.

## 2.6 Data Analysis Methods

Here we review some data analysis techniques relevant to this thesis that do not fit specifically into the category of chemometrics.

### 2.6.1 Correlation

The correlation between two variables  $x_i$  and  $y_i$  is a measure of the quality of the linear fit. The coefficient of variation  $R^2$  between two variables is defined as (Ott and Mendenhall, 1985)

$$R^2 = \frac{SS_{xy}^2}{SS_{xx}SS_{yy}}, \quad (2.44)$$

where

$$\begin{aligned} SS_{xy} &= \sum_i (x_i - \bar{x})(y_i - \bar{y}), \\ SS_{xx} &= \sum_i (x_i - \bar{x})(x_i - \bar{x}), \\ SS_{yy} &= \sum_i (y_i - \bar{y})(y_i - \bar{y}), \end{aligned} \quad (2.45)$$

are the sum of squares about the mean ( $\bar{x}$  and  $\bar{y}$ ). If  $y$  is modelled from  $x$  then  $R^2$  describes the proportion of variation in  $y$  that is explained by  $x$  (Ott and Mendenhall, 1985).

Assuming a linear fit exists,  $R^2 = 1$  means a perfect straight line fit and  $R^2 < 1$  means there is some deviation from the line. Nonlinearity in the relationship between two variables reduces  $R^2$  when computed according to equation 2.44, even when a strong relationship exists. It can be necessary to visually inspect the scatter of  $x$  verses  $y$  for the presence of nonlinearity.

When considering the correlation between two univariate variables (or a univariate with each variate in a multivariable), then we write  $r$ , which is computed as the square root of equation 2.44. The term  $r$  is known as the correlation coefficient. Unlike  $R^2$ ,  $r$  may be negative, where negative values indicate negative proportionality.

### 2.6.2 Variography

Variography (Gy, 2004a,b; Petersen and Esbensen, 2005) is a tool from the theory of sampling (TOS) for examining variation in measurements over some dimension. Consider a set of objects to be measured called units. The population of units is called the lot  $L$ . The heterogeneity  $h_m$  is the “contribution to the heterogeneity of lot  $L$  made by unit  $U_m$ ” (Gy, 2004a). Given measurements,  $a_m$  on objects with mass  $M_m$ , the heterogeneity is

$$h_m = N_U \frac{(a_m - a_L) M_m}{a_L M_L} \quad (2.46)$$

where  $a_L$  and  $M_L$  are the lot measurement and mass respectively, and  $N_U$  is the number of units. Given a lag parameter,  $j$ , representing a physical gap between samples over the dimension of interest the variogram is computed as

$$V(j) = \frac{1}{2(N_U - j)} \sum_m (h_{m+j} - h_m)^2. \quad (2.47)$$

Lag can be any physical parameter, such as distance, time or even just the count of units going past on a conveyor belt. The variogram represents the spatial correlation against distance informing of the nature of variability with increasing distance from any given sample. The variogram can be decomposed into four parts: random fluctuations, the non-random continuous part, the non-random continuous cyclic part and some residual part. The random part is due to random fluctuations within each unit. The non-random continuous part describes the variation trend with lag, e.g. if variability increases, decreases or is static with distance. Similarly the cyclic part is due to cyclic variation. The residual, typically small, is due to error in estimation.

### 2.6.3 Receiver Operator Characteristics and the Wilcoxon

Consider a two class detection problem where there are positive instances (or measurements) and negative instances. Now if an instance is truly positive, as identified by some reference method, then ideally we want a classifier to determine that measurement to be positive and likewise for negative instances. Such classifications are called true positives and true negatives. When a classifier is in error and labels a negative instance positive we have a false positive. Similarly misclassified positive instances are called false negatives. Let  $N_p$  be the number of positive instances and  $N_n$  the number of negative instances. Thus the total number of instances is  $N = N_p + N_n$ . Given the number of false positives, FP, made by a classification and the number of true positives, TP, then the true positive

fraction, TPF, and false positive fraction, FPF, are

$$\text{TPF} = \frac{\text{TP}}{N_p}, \text{FPF} = \frac{\text{FP}}{N_n}. \quad (2.48)$$

Frequently classification algorithms have a continuous valued output to which some threshold parameter is applied to delineate positive and negative classifications. By varying this threshold parameter a set of TPFs and FPFs are observed. By plotting the TPF against  $1 - \text{FPF}$  a receiver operator characteristic (ROC) curve is obtained. The area under the ROC curve (AUC) is indicative of the power of the classification method, with an AUC of 1 being perfect classification and an AUC of 0.5 no better than random classification. To compute the AUC numerical integration of the ROC can be used (Hanley and McNeil, 1982), the so-called non-parametric AUC.

The AUC is equivalent to the Wilcoxon statistic (Hanley and McNeil, 1982). For nonparametric data of two groups, count the number of times each member of one group is larger than each member of the other group. The sum of the counts is indicative of the separation between the groups. Division of the sum of counts by the product of the number of members in the two groups,  $N_p N_n$ , gives the Wilcoxon. Like the AUC the Wilcoxon is a fraction between 0.5 and 1. Computing the Wilcoxon on the intermediate continuous classification output with group assignments according to the reference method yields the AUC. Cortes and Mohri (2004) observed that for two groups, A and B, the AUC is the probability that A is greater than B, that is  $P(A > B)$ , and that the Wilcoxon is the same for the discrete case. Thus they proved the link between the AUC and the Wilcoxon.

The null hypothesis,  $H_0$ , for the Wilcoxon is that there is no separation between the groups,  $H_1$  is that there is separation. The probability that random group assignment yields a larger Wilcoxon than the reference determines the significance. If the probability is sufficiently low then  $H_0$  is rejected in favour of  $H_1$ . Estimation of the probability is performed by repeatedly assigning random grouping to the data and computing the Wilcoxon, say one hundred or one thousand times. The p-value is the fraction of times that a random Wilcoxon is greater than the value due to the reference.

## 2.7 Near Infrared Spectroscopy and Beef

For over a decade beef grading based on NIR measurements has been a topic of interest. Traditional measurement of chemical content, such as fat, water or acidity or functional features such hardness (first bite) and tenderness (relating to the whole chewing process) typically required slow and destructive techniques. NIR on the other hand is fast and nondestructive, so if mathematical models relating spectra of meat to relevant features are accurate then spectroscopy holds significant potential to perform real time online grad-

ing. Tenderness is assessed two ways, by expert panel and by shear force. Shear force measurements are taken using a metal ‘tooth’ that bites through the meat with a sensor that measures the amount of force required. The majority of tenderness readings in the literature are taken with a Warner-Bratzler shear force meter. A recent review by Huang *et al.* (2008) covers the prominent literature on meat features predicted or analysed via NIR spectroscopy. In reviewing, we convert all correlation coefficients reported as  $r$  to  $R^2$  for consistency. Where available and relevant, root mean square error in crossvalidation (RMSECV), root mean square error in prediction (RMSEP), and the standard deviation of the reference data (reference SD) are included.

Hildrum *et al.* (1994) tested prediction of hardness, tenderness and juiciness of beef longissimus dorsi (strip loin) muscles by NIR spectroscopy. They present spectra in the band 1100–2500 nm with 4 nm intervals and also consider transmission spectra in the band 850–1050 nm with 2 nm intervals. The reference variables were determined by an expert panel of nine trained assessors. The hardness ranged from 2.9–7.1, the tenderness ranged from 2.9–7.4 and the juiciness ranged from 4.6–6.1 (each reference value was on a scale from 1 to 9.) They used PCR with MSC preprocessing of the spectra. Using cross validation they obtained correlation for hardness and tenderness with  $R^2$  in the range 0.64–0.81, with reported root mean square error of cross validation (RMSECV) 0.54–1.04, but did not obtain good correlation for juiciness ( $R^2$  in the range 0.13–0.24, RMSECV 0.36–0.51). The authors make interesting comments on how tougher samples have higher absorptions than tender samples at most wavelengths and that a the spectrum of a tender sample is affected more by the freezing than a tougher sample, but do not seek a spectroscopic basis to the models they build. Rather they discuss in detail the data analysis, considering carefully the process of model building and outlier removal.

Naes and Hildrum (1997) examined classification methods for tenderness prediction by NIR spectroscopy and compared classification to regression on their data. They acquired spectra between 1100–2500 nm on raw meat samples. Based on an expert panel assessment of the meat, they grouped the samples into three tenderness groups: very tough, intermediate and very tender. They used a variety of methods, including: grouping of PCR output; canonical discriminant analysis; Bayes rule and variations thereof including the Mahalanobis distance on the principal components. Using classification they obtained good results on extreme tenderness values. On the other hand PCR calibration tended to return intermediate values, so samples that were very tender or very tough were often misclassified by examination of the regression result. They concluded that classification was better than PCR for tenderness prediction and that the Mahalanobis distance was the best of the classification methods used. The Mahalanobis distance utilises information about the mean and distribution of the different data groups, classifying data according to what distribution each point fits best into. Later in this thesis we utilise a

related technique to perform classification.

Andersen *et al.* (1999) examined pH with visible (362–777 nm, 1 nm intervals) and NIR (1000–2630 nm, 1 nm intervals) spectroscopy in pork. They examined correlations per wavelength and calibration using PLS. The pH measurements were taken of 46 samples the day after slaughter. They compared the pH of the meat while whole to after homogenisation and found that they correlate well ( $R^2 = 0.978$ ). On the unhomogenised samples they had pH values in the range 5.46–6.97. Correlations of the pH with the individual wavelengths varied widely from a positive correlation with  $R^2 = 0.64$  at 360 nm to a negative correlation with  $R^2 = 0.30$  at 1900 nm. PLS modelling of the pH with the spectra was performed. Leave one out crossvalidation was used, from which they obtained a correlation of  $R^2 = 0.53$  over LD muscles with the spectral band 1000–2630 nm (reference SD 0.153, RMSECV 0.104).

Rødbotten *et al.* (2000) examined the relationship between spectra (1100–2500 nm, 4 nm intervals) taken early post-mortem and quality features. They examined final tenderness (both tenderometer and sensory) and fat content using PLS with and without MSC using full crossvalidation. They observed moderately poor prediction of tenderometer readings with  $R^2$  values of up to 0.46 and worse prediction of sensory values with  $R^2$  in the range 0.22–0.30. Applying MSC to the spectra did not improve the correlation with tenderness. Moderate correlation with intramuscular fat content (0.61–0.72, RMSECV 1.2–1.4 %) was observed. They concluded that their findings do not support the use of spectra taken early postmortem for final tenderness prediction, but stated that broadening the spectral bandwidth may improve the prediction result.

Rødbotten *et al.* (2001) measured NIR spectra (950–1700 nm, 6 nm intervals) on beef at 2, 9 and 21 days post mortem with two treatments based on storage temperature to affect tenderness. The tenderness ranges at the respective days post-mortem were 5.94–14.8 kg cm<sup>-2</sup>, 3.26–12.2 kg cm<sup>-2</sup> and 3.04–15.1 kg cm<sup>-2</sup>. PLS was used to perform tenderness prediction from the spectra. Comparing spectra with the tenderness readings taken at the same time, including the temperature treatment information for calibration did not improve the power of the spectroscopy to predict tenderness. For prediction of tenderometry readings in the future, including the treatment information assisted in tenderometry prediction. Crossvalidation was used to assess prediction power. Calibration without treatment information gave prediction with  $R^2$  in the range 0.27–0.69 (RMSECV in the range 2.54–1.37 kg cm<sup>-2</sup>). Including the treatment information in the calibration improved correlation to 0.50–0.72 (RMSECV in the range 2.15–1.29 kg cm<sup>-2</sup>). The pooled SD of error in the reference ranged from 0.48 kg cm<sup>-2</sup> at 2 days to 0.62 kg cm<sup>-2</sup> at 21 days. Also measurement of spectra through a glass plate improved predictions, which the authors tentatively ascribe to reduced specular reflection on the meat surface but suggest further examination before final conclusions are drawn. The beef samples were as-

signed to one of two or three groups by their tenderometry readings. When two groups were used 73–98 % of the samples were correctly classified, when three groups were used 63–75 % were correctly classified. It may be that the spectral band improved results over the earlier study, but it is difficult to decouple any improvement due to the spectral band from the improvement due to the more detailed timepoint information acquired. A commonly reported feature in the literature reviewed is a trend for overall NIR reflectance to decrease with tenderness.

Tøgersen *et al.* (2003) examined fat, moisture and protein in ground beef by NIR spectroscopy at several discrete temperatures between  $-5$  and  $+10$  °C. The reference value ranges were, in percentage of wet weight, fat 7.66–22.91 %, moisture 59.39–71.48 %, and protein 17.04–20.76 %. The spectrometer measured the band 1100–2500 nm with 2 nm intervals. Principal component regression was used on wavelength bands selected to avoid the water freezing effect on the spectra. The reference measurements were obtained by wet chemistry. Good prediction results, presented as the root mean squared error of cross validation (RMSECV) percentage of wet weight, were obtained in each case. The RMSECV were 0.48–1.11 % for fat, 0.43–0.97 % for moisture and 0.41–0.47 % for protein.

González-Martín *et al.* (2003) examined a wide range of specific subcutaneous fatty acids in Iberian swine using NIR in the range 1100–2000 nm with resolution of 2 nm. They used modified PLS with several preprocessing measures (SNV, MSC, detrending and numerical differentiation), selecting the preprocessing that gave the best results under crossvalidation. An external test set was used for validation, where the authors indicate that the external data were acquired on new samples. Poor to good  $R^2$  and standard error in prediction values were reported. In particular the  $R^2$  values ranged from 0.17 to 0.94. Overall they conclude that "...the NIRS technique, using a fibre-optic probe, is a useful alternative to gas chromatography for determination of fatty acid contents in samples of subcutaneous fat from Iberian pigs." Sierra *et al.* (2008) used near infrared transmittance (NIT) in the range 850–1050 nm, 2 nm intervals, for a range of specific fatty acids, similar to that of González-Martín *et al.* (2003), obtaining good prediction of some prominent species. They also observed accurate prediction of groups such as saturated ( $R^2 = 0.837$ ), branched ( $R^2 = 0.701$ ) and monounsaturated ( $R^2 = 0.852$ ) fatty acids.

Qiao *et al.* (2007a) used a pushbroom hyperspectral imager to examine pork for marbling levels and quality features, such as redness or paleness and exudation. The imager was a CMOS camera based linescan system that acquired spectra in the range 400–1000 nm with spectral resolution 2.8 nm. Principal components of the spectra were computed and subjected to cluster analysis and a feed forward neural network. The reddish firm samples were grouped into exudative and non-exudative groups with 75–80 % success. An angular second moment was used as a texture feature, and applied to determine

marbling scores. The same imaging system was used to examine for colour, pH and drip loss (Qiao *et al.*, 2007b). Drip loss was determined as the percentage of weight loss after two days storage and four degrees. Colour was the luminance, L, of La\*b\* space taken with a chromameter. Wavebands were selected by choosing those that best correlated with the response variables. Prediction was performed by taking the “average intensity of whole loin-eye area at each feature band images.” The drip loss was predicted with a correlation coefficient ( $R^2$  value) of 0.59, the pH 0.30 and the colour, L, 0.74.

Andrés *et al.* (2008) examined NIR over 400–2498 nm, 2 nm intervals, for prediction of pH, sarcomere length, cooking loss, shear force and colour (La\*b\*) in beef at 1, 3, 7 and 14 days post slaughter. They took spectra at the same time points as observing reference features. Using PLS, good predictability was observed for pH at 24 hours ( $R^2 = 0.97$ , Range 5.50–6.67, reference SD = 0.33) and luminance (L) at cut time ( $R^2 = 0.85$ , Range 27.62–42.70, reference SD = 2.76) and 60 minutes ( $R^2 = 0.82$ , Range 28.89–43.78, reference SD = 3.76). The rest of the examined features had poorer parameters.

Scattering coefficients computed on beef muscle were computed and examined with respect to tenderness (Xia *et al.*, 2007) and heat induced properties (Xia *et al.*, 2008). The scattering coefficients were derived from a physical theoretical model of oblique-incidence reflectometry. The model includes consideration of effective attenuation, angle of incidence and the lateral offset of the points of incident light to the centre of the diffuse reflectance. From the model they obtained an absorption coefficient and a ‘reduced scattering’ coefficient. “These absorption and scattering coefficients represent the probabilities of a photon being absorbed and scattered inside the sample” (Xia *et al.*, 2007). Light from an optical fibre conduit source impinged on the meat sample at an oblique angle, and a collection fibre at a right angle to the meat conducted reflected light to a spectrometer. The sampling fibre was moved laterally to obtain a profile of reflectance with position. An optical scattering model was fit to the profile at each wavelength, from which the absorption and scattering coefficients were computed. The scattering coefficient at 721nm was found to correlate with the shear force with  $R^2 = 0.59$  (precise statistics of the reference data are not given, but from the figures presented we surmise a shear force range of about 29–73 N). They recognise that the result, while encouraging, requires further study with large samples sizes (they used 32) and the use of multivariate statistics to explore the relationship of tenderometry readings with the whole spectrum. Scattering coefficients were observed to change with the structural changes caused by heating. Also, based on their observations of how tenderometry readings and scattering coefficients change with temperature, they deduced that myofibril (muscle fibres that run across the cell) and collagen (connective tissue) have different effects from each other on scattering. Overall the work of Xia *et al.* (2007, 2008) highlights that scattering effects are relevant when considering optical measurements on meat.

Bowling *et al.* (2009) used spectroscopy in the band 345–1100 nm, 5 nm, to predict beef tenderness. They used linear regression, over averages of spectral bandwidths, regression over band averages with product and squared terms (second order polynomial) over bandwidths in the visible domain and PCR. The terms that entered the regression model were chosen using a forward selection procedure (very little detail is given). They reported low correlations ( $R^2 \leq 0.23$ ,  $RMSE \geq 0.754$ , rather than use independent set validation, they test for significance using regression analysis) with the best correlation occurring for the regression including the squared terms. They observed that using the visible and the NIR together did not improve the predictive ability of the models. They noted that other authors had reported high correlations, but those prior studies used greater variation in muscles and treatment types. Rosenvold *et al.* (2009) examined the relationship between NIR spectroscopy and quality features. They examine the spectral band 400–1700 nm with wavelength spacing 6 nm. Several treatments were used to create a range of values in the measurements, such as electrical stimulation, wrapping of the muscles and varying the cooling temperature. They measure features pre-rigor such as pH, glycogen and temperature and and post rigor like pH, shear force using a MIRINZ tenderometer, (MacFarlane and Marer, 1966) and water holding capacity. The spectra and reference measurements were taken at intervals post-slaughter to observe changes over time. They observed that the first principle component of the spectra separated the spectra into pre-rigor and post-rigor groups. Calibrations were built using PLS, the predictive power of which was assessed using independent set validation. For pre and post-rigor pH they obtained  $R^2 = 0.84$  (range 5.15–7.17, RMSEP 0.2), for post-rigor tenderness (shear force)  $R^2 = 0.58$  (range 19–265 N, RMSEP 28 N), and for pre-rigor glycogen  $R^2 = 0.70$  (range 0.0–18.7 mg/g, RMSEP 2.7 mg/g). They also obtained a calibration of the natural logarithm of time in hours post-rigor with  $R^2 = 0.82$  (range 0.0–4.5  $\ln(\text{hours} + 1)$ ), RMSEP 0.66  $\ln(\text{hours} + 1)$ ).

The spatial variation of the textural and shear force properties in meat has been studied (Hansen *et al.*, 2004; Janz *et al.*, 2006). Shear, tenderness, hardness, juiciness and sensory textural attributes were all found to vary within different muscles. Significant sensory and instrumental differences between the left and right side of the carcass were observed.



# Chapter 3

## Theory

This chapter covers the theory for optical multiplexing and extends the theory where necessary. Hadamard multiplexing, the connection of Hadamard multiplexing with strongly regular graphs and compressed sensing are covered.

The action of Hadamard transform matrices on realistic types of error is expounded. In covering the Hadamard multiplexing we develop a unified notational framework under which the action of the different matrices are computed and compared. Broadly the error is divided into two groups, error that does not interact with the encoding at encoding (additive error) and error that does (multiplicative error). Furthermore random and systematic error are both considered for each group. The theory for the correction of the systematic error effects at decoding is developed and the total combined effect is examined.

A link between the Hadamard transform S-matrix matrix and a specific type of matrix from graph theory was described by Ratner *et al.* (2007) via the respective eigenvalue structure and subsequent action on noise by multiplexing. Here the link is explored further for the other Hadamard matrices. The construction of the S-matrix from the H-matrix is reversed for the graph theory matrix. The A-optimality (average optimality) of the resultant G-like and H-like matrices is examined via the eigenvalue structure and compared to the Hadamard equivalents.

The recently developed method of compressed sensing falls in the realm of multiplexing. If a digital signal to be reconstructed has  $N$  data points then, unlike traditional multiplexing, compressed sensing involves taking fewer than  $N$  measurements. Hence the acquisition time is shorter than traditional multiplexing. The reconstruction is a compressed version of the sensed signal. The accuracy of compressed sensing, that is the boundedness of the error due to taking fewer measurements, is well established. Here the effect on the precision due to the number of compressed sensing measurements is developed. The acquisition is typically performed using a matrix of randomly assigned values. The A-optimality is derived by inference using known eigenvalue properties of large random rectangular matrices.

### 3.1 Multiplexing

A multiplexing system is one where data is acquired according to some linear transform  $M$ , viz

$$\mathbf{a} = M\mathbf{p}, \quad (3.1)$$

where  $\mathbf{p}$  is the unknown data of interest and  $\mathbf{a}$  is the data physically acquired. We acquire  $\mathbf{a}$  according to  $M$  and then reconstruct  $\mathbf{p}$  by inversion of the transformation, namely

$$\mathbf{p} = M^{-1}\mathbf{a}. \quad (3.2)$$

The acquisition in Eq 3.1 must be invertible either by direct matrix inversion of  $M$  to recover  $\mathbf{p}$ , or the estimation of  $\mathbf{p}$  must be computable by more sophisticated algorithmic means.

Due to practicalities, the entries in the matrix  $M$  are typically bound to  $\{M\}_{ij} \in [-1, 1]$ . Physically the bound on the entries in  $M$  means that each ‘datum of interest’ in  $\mathbf{p}$  contributes at most once to each measurement in  $\mathbf{a}$ . In certain cases the entries of  $M$  may range continuously between the bounds, such as Fourier transform spectroscopy (Becker, 1972) or magnetic resonance imaging (Westbrook, 2002) where  $M$  is the Fourier transform matrix. In contrast for multislit spectroscopy (Harwit and Sloane, 1979) or optically multiplexed imaging (the topic herein) each entry in  $M$  must be one of the set  $\{-1, 0, 1\}$ . Each row in  $M$  corresponds to one acquisition and each column one datum. For the  $j$ th row in  $M$  an  $i$ th column entry of  $-1$ ,  $0$  or  $1$  means that the  $i$ th datum must be subtracted, excluded or added respectively to obtain the corresponding  $j$ th entry in  $\mathbf{a}$ . The case  $M = I$  ( $I$  the identity matrix) corresponds to setting  $j = i$  and physically means measuring each entry in  $\mathbf{p}$  individually, which we refer to as pointwise acquisition.

#### 3.1.1 Hadamard Matrices

Hadamard multiplexing is the acquisition of data according to one of the Hadamard matrices (Harwit and Sloane, 1979). The Hadamard matrices are the H-matrix, S-matrix and G-matrix. The H-matrix,  $H_N$ , has the maximum possible determinant for any  $N \times N$  matrix with entries constrained between  $-1$  and  $1$ . Every entry in  $H_N$  has absolute value  $1$  and the H-matrix satisfies

$$\begin{aligned} H_N^T &= H_N, \\ H_N^T H_N &= H_N^2 = N I_N, \\ \Rightarrow H_N^{-1} &= \frac{1}{N} H_N, \end{aligned} \quad (3.3)$$

where  $\cdot^T$  is the matrix transpose. The self transpose condition is not strictly necessary, but is useful for notational simplicity and ease of computation in the theoretical analysis.

The Sylvester H-matrix construction is as follows (Harwit and Sloane, 1979). The matrices  $H_1$  and  $H_2$  are:

$$\begin{aligned} H_1 &= [1], \\ H_2 &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \end{aligned} \quad (3.4)$$

Given two H-matrices,  $H_N$  and  $H_M$ , a new H-matrix  $H_{NM}$  is constructed by replacing each entry in  $H_N$ ,  $\{H_N\}_{i,j}$ , with  $\{H_N\}_{i,j} H_M$ , that is by taking the Kronecker product between  $H_M$  and  $H_N$ . This construction gives matrices of the order  $N = 2^k$ ,  $k = 1, 2, 3, \dots$ . It is conjectured that Hadamard matrices exist for all values of  $k$ .

The G-matrix,  $G_{N-1}$ , is obtained from the H-matrix by removing the leading row and column of ones, hence

$$H_N = \begin{bmatrix} 1 & \mathbf{o}_{N-1}^T \\ \mathbf{o}_{N-1} & G_{N-1} \end{bmatrix}, \quad (3.5)$$

where  $\mathbf{o}_{N-1}$  is a column vector of ones of length  $N - 1$ . The S-matrix is subsequently derived from the G-matrix by changing all the 1s to 0 and the  $-1$ s to 1. Clearly,

$$G_{N-1} = J_{N-1} - 2S_{N-1}, \quad (3.6)$$

where  $J_{N-1}$  is a matrix of ones. From the fact (Harwit and Sloane, 1979),

$$\begin{aligned} G_{N-1}S_{N-1} &= J_{N-1}S_{N-1} - 2S_{N-1}^2 \\ &= \frac{N}{2}J_{N-1} - \frac{N}{2}(I_{N-1} + J_{N-1}) \\ &= -\frac{N}{2}I_{N-1}, \end{aligned} \quad (3.7)$$

the inverses of  $S$  and  $G$  are,

$$\begin{aligned} S_{N-1}^{-1} &= -\frac{2}{N}G_{N-1}, \\ G_{N-1}^{-1} &= -\frac{2}{N}S_{N-1}. \end{aligned} \quad (3.8)$$

An alternative direct construction of S-matrices is via maximal length shift-register sequences (SRS) (Harwit and Sloane, 1979). A SRS S-matrix is cyclic (each row is a cyclically shifted version of the previous one) of order  $N - 1 = 2^m - 1$ ,  $m = 1, 2, 3, \dots$ . To describe the construction process we follow the example given by Harwit and Sloane (1979).

1. Take a binary primitive polynomial, for example  $x^4 + x + 1$  which corresponds to  $m = 4$ .
2. Set the entries of the first row of  $S$  to  $\{S\}_{1,1} = 1$  and  $\{S\}_{1,j} = 0, j \neq 1$ .
3. Cyclically shift each element in the first row of  $S$  one step to the right and at each entry take the sum modulo 2 according to the binary primitive polynomial, that is,  $\{S\}_{1,j+4} = \{S\}_{1,j+1} + \{S\}_{1,j}$ . Repeat until no further change occurs.
4. For the rows from  $i = 2$ , iteratively set  $\{S\}_{i,j} = \{S\}_{i-1,j+1}$ .

A list of binary primitive polynomials for  $m = 1, \dots, 20$  is given by Harwit and Sloane (1979). Other methods for constructing cyclic S-matrices exist, a review of such methods is also given by Harwit and Sloane (1979).

### 3.2 Error Effects Under Hadamard Encoding

In practical multiplexing the acquired data are contaminated by random and systematic error. The random component due to noise causes a variance in measurement; the systematic component causes a bias. Considering noise, the primary purpose of multiplexing is to permit more of the object being measured (light in our case) to the measurement device so that the signal is increased. The signal increase causes a signal to noise ratio (SNR) boost at data acquisition. If the multiplexing pattern is optimal then the SNR boost is optimally maintained at reconstruction. This section addresses general key forms of random noise and systematic error in multiplexing. The contributions to the state of the art of this section are:

- A coherent and unified description of the effect of all likely error sources for all three types of Hadamard matrix.
- An examination of drift *and* Poisson noise effects for all the Hadamard matrices and, in particular, for the H-matrix and G-matrix.

The error types and their effect under multiplexing are introduced individually. The error types are: random additive noise, additive offset, Poisson noise or any noise with variance proportional to the signal, and systematic multiplicative offset. The correction of the systematic error types and the subsequent effect on the random noise are considered. This section does not consider the physical cause of the error types explicitly to maintain generality. While the theory was developed in the context of optical multiplexing, applicability to other multiplexing situations is possible by reconsideration of the error sources.

### 3.2.1 Additive Random Noise

We first consider the effect of random additive error,  $\mathbf{e}$ , with variance  $\sigma^2$ , that occurs at acquisition and is independent of the multiplexing matrix  $M$

$$\mathbf{a} = M\mathbf{p} + \mathbf{e}. \quad (3.9)$$

Decoding recovers an estimate of the data,  $\mathbf{p}$ , contaminated with additive error that is modified by the inverse of the encoding system

$$\begin{aligned} \hat{\mathbf{p}} &= M^{-1}\mathbf{a}, \\ &= M^{-1}M\mathbf{p} + M^{-1}\mathbf{e}, \\ &= \mathbf{p} + \boldsymbol{\epsilon}, \end{aligned} \quad (3.10)$$

where  $\boldsymbol{\epsilon} = M^{-1}\mathbf{e}$ .

An advantageous encoding system will reduce the variance of the error at decoding. The variance (or MSE) of the errors at encoding are not guaranteed to all be the same as each other at decoding, nor are the MSE of the errors at decoding. To assess the theoretical performance of multiplexing system matrices we use the average of the MSE or A-optimality (Harwit and Sloane, 1979). Other optimality measures exist, but A-optimality is probably the most important measure provided that the errors are approximately equal (Harwit and Sloane, 1979). Let  $\epsilon_i$  be the entries in the vector  $\boldsymbol{\epsilon}$ , an A-optimal matrix minimises

$$\epsilon = \frac{1}{N} \sum_{i=1}^N \epsilon_i^2. \quad (3.11)$$

It turns out that an A-optimal design is one which minimises (Harwit and Sloane, 1979)

$$\epsilon = \frac{\sigma^2}{N} \text{Tr} [(M^T M)^{-1}], \quad (3.12)$$

where  $\text{Tr}[\cdot]$  is the matrix trace. The trace term in equation 3.12 is useful for analysing the predicted error behaviour of a matrix  $M$  for random additive noise.

For pointwise acquisition,  $M = I$ , we have

$$\text{Tr} [(I_N^T I_N)^{-1}] = \text{Tr} [I] = N. \quad (3.13)$$

Equation 3.13 is the baseline for multiplexing, any multiplexing matrix that does not yield a smaller result provides no advantage.

For the H-matrix we have using equation 3.3

$$\begin{aligned}\text{Tr} [(H_N^T H_N)^{-1}] &= \frac{1}{N^2} \text{Tr} [H_N^2] \\ &= \frac{1}{N^2} N \text{Tr} [I_N] \\ &= 1.\end{aligned}\tag{3.14}$$

The H-matrix attains the smallest possible value according to Hotelling's bound on the MSE (Harwit and Sloane, 1979), thus the H-matrix is optimal for multiplexing. The optimality is also seen intuitively by noting that the inverse of  $H$  is simply a scaled version of itself, that is, the rows in  $H$  form an orthogonal basis set. The reconstruction then is essentially a rotation and scaling operation and the reconstructed data have the same SNR as the acquired data on average. If the SNR is optimised at acquisition, then the MSE is minimised at reconstruction.

For the G-matrix we have

$$\begin{aligned}\text{Tr} [(G_{N-1}^T G_{N-1})^{-1}] &= \text{Tr} [G_{N-1}^{-2}] \\ &= \frac{4}{N^2} \text{Tr} [S_{N-1}^2]\end{aligned}\tag{3.15}$$

where equation 3.8 was used. An S-matrix has on each row and column  $(N - 1 \pm 1)/2$  ones, thus  $\text{Tr}(S_{N-1}^2) = (N - 1)(N - 1 \pm 1)/2$ . The  $\pm 1$  term for  $S$  encompasses both Sylvester and cyclic SRS matrices, with the former corresponding to  $-1$  and the latter to  $+1$ . Now

$$\begin{aligned}\frac{4}{N^2} \text{Tr} [S_{N-1}^2] &= \frac{4}{N^2} \frac{(N - 1)(N - 1 \pm 1)}{2}, \\ &\approx 2, \text{ for large } N.\end{aligned}\tag{3.16}$$

So  $G$  is not as optimal as  $H$  but provides a useful alternative should a matrix of odd dimensions be required. This result was presented by Sloane *et al.* (1969) and was independently derived by the present author (Streeter *et al.*, 2008b) without knowledge of the derivation by Sloane *et al.* (1969).

For the S-matrix we rework the result of (Sloane *et al.*, 1969; Harwit and Sloane, 1979) in the manner above to obtain

$$\begin{aligned}\text{Tr} [(S_{N-1}^T S_{N-1})^{-1}] &= \text{Tr} [S_{N-1}^{-2}], \\ &= \frac{4}{N^2} \text{Tr} [G_{N-1}^2].\end{aligned}\tag{3.17}$$

Since the  $G$ -matrix has no zeros we see that  $\text{Tr} [G_{N-1}^2] = (N - 1)^2$ , which gives

$$\begin{aligned} \frac{4}{N^2} \text{Tr} [G_{N-1}^2] &= \frac{4}{N^2} (N - 1)^2, \\ &\approx 4, \text{ for large } N. \end{aligned} \quad (3.18)$$

The  $S$ -matrix, while not as optimal as the  $H$ -matrix or the  $G$ -matrix, contains no negative terms. The  $S$ -matrix is conjectured to be the optimal matrix with no negative terms (Harwit and Sloane, 1979) and is useful in situations where the subtractions necessary for implementing  $H$  and  $G$  are not practicable. As such the  $S$ -matrix is very common in multiplexing.

### 3.2.2 Additive Offset Bias

Additive offsets cause bias in the measurement. Systematic additive components such as background and baseline offset are represented by a term that is multiplexed,  $\mathbf{t}_1$ , and a term that is independent of the multiplexing  $\mathbf{t}_2$ . Multiplexed acquisition in the presence of such offsets is

$$\mathbf{a} = M(\mathbf{p} + \mathbf{t}_1) + \mathbf{t}_2. \quad (3.19)$$

Reconstruction by inversion of  $M$  gives

$$\hat{\mathbf{p}} = M^{-1} \mathbf{a} = \mathbf{p} + \mathbf{t}_1 + M^{-1} \mathbf{t}_2. \quad (3.20)$$

The independent component,  $\mathbf{t}_2$ , is reduced by multiplexing whereas  $\mathbf{t}_1$  is recovered.

Being systematic the additive offsets must be directly measured and subtracted at the appropriate point in reconstruction. Measuring  $\mathbf{t}_2$  is performed directly by setting  $M = 0$ , generally a simple matter in many multiplexing systems. The multiplexed offset is more problematic because 3.19 implies that  $\mathbf{t}_1$  cannot be measured independently of  $\mathbf{p}$ . An auxiliary independent measurement device observing the measurement environment may overcome the difficulty in measuring  $\mathbf{t}_1$ , but also introduce new error sources. Thus it is preferable to avoid multiplexing systems for which  $\mathbf{t}_1 \neq 0$ .

Additive offset factors that drift are represented by varying entries in  $\mathbf{t}_1$  and  $\mathbf{t}_2$ . When the offset is guaranteed not to drift then only one measurement is required which is subsequently subtracted appropriately during reconstruction. Prevention of drift altogether is very difficult, if not impossible, and typically the entries  $t_{1,j}$  and  $t_{2,j}$  in  $\mathbf{t}_1$  and  $\mathbf{t}_2$  must be measured with each  $a_j$  in  $\mathbf{a}$ . Making such measurements gives estimates of the true offset

values with random error  $\hat{\mathbf{t}}_1$  and  $\hat{\mathbf{t}}_2$ . The measurements taken are modelled as

$$\begin{aligned}\mathbf{a} &= M(\mathbf{p} + \mathbf{t}_1) + \mathbf{t}_2 + \mathbf{e}_a, \\ \hat{\mathbf{t}}_1 &= \mathbf{t}_1 + \mathbf{e}_1, \\ \hat{\mathbf{t}}_2 &= \mathbf{t}_2 + \mathbf{e}_2,\end{aligned}\tag{3.21}$$

where  $\mathbf{e}_a$ ,  $\mathbf{e}_1$  and  $\mathbf{e}_2$  are sampled at random from the same distribution<sup>1</sup>. Making the appropriate corrections

$$\hat{\mathbf{p}} = M^{-1}(\mathbf{a} - \hat{\mathbf{t}}_2) - \hat{\mathbf{t}}_1,\tag{3.22}$$

we see that

$$\hat{\mathbf{p}} = \mathbf{p} - \mathbf{e}_1 + M^{-1}(\mathbf{e}_a - \mathbf{e}_2).\tag{3.23}$$

Thus from equation 3.23 correction of  $\mathbf{t}_1$  and  $\mathbf{t}_2$  at reconstruction results in an average MSE of

$$\epsilon = \sigma^2 + \frac{2\sigma^2}{N} \text{Tr} [(M^T M)^{-1}].\tag{3.24}$$

Correction of  $\mathbf{t}_2$  increases the average MSE of the error that is reduced by multiplexing. However correction of  $\mathbf{t}_1$  introduces error that is not affected by the multiplexing, which reinforces the preference of avoiding situations where  $\mathbf{t}_1 \neq 0$ .

### 3.2.3 Poisson Noise

The variance of Poisson noise is proportional to the mean. Multiplexing, which boosts the signal at data acquisition, also increases the variance of Poisson noise. The critical issue for multiplexing in the presence of Poisson noise is how much the encoding increases the noise at the data acquisition compared to how much the reconstruction reduces the noise. We first derive the general form of the average MSE of decoded multiplexed estimates in the presence of Poisson noise, then examine the specific Hadamard matrices in turn. We restrict the analysis to the case where signal proportional noise emanates from sources that add positively, so the Poisson noise variance is therefore as large as it can be so the following analysis represents a ‘worse case scenario.’ Conceptually there may be cases where the noise is proportional to the signal reduced by the negative encoding elements in  $H$  and  $G$ , causing a smaller noise variance than in the following analysis.

Let  $\langle r \rangle$  be an average intensity ‘weight’ in measurement and  $E$  be a diagonal matrix with random values to represent noise. Furthermore let each diagonal element in  $E$  be sampled at random from a distribution with variance  $\langle r \rangle$ . In the context of reflectance imaging or spectroscopy  $\langle r \rangle$  is the mean intensity of the light source and is typically a

<sup>1</sup>If  $\hat{\mathbf{t}}_1$  is taken with an axillary device then  $\mathbf{e}_1$  is from a different distribution.

large real number (assuming a classical model of the physics). In other contexts it may be physically sensible for  $\langle r \rangle$  to have any other magnitude or even be unity. We model multiplexing with signal dependent noise as

$$\mathbf{a} = \langle r \rangle M \mathbf{p} + E M \mathbf{p}. \quad (3.25)$$

So an error term is multiplied by *each encoded measurement* and added to the noise free encoding. Decoding gives

$$\hat{\mathbf{p}} = \langle r \rangle \mathbf{p} + M^{-1} E M \mathbf{p}. \quad (3.26)$$

Let  $\mathbf{e}_{P,\mathbf{a}} = E M \mathbf{p}$ , the noise at acquisition. The variance in the  $j$ th entry of the acquisition  $\mathbf{a}$ ,  $\sigma_{P,j,\mathbf{a}}^2$ , due to the  $j$ th entry in  $\mathbf{e}_{P,\mathbf{a}}$  is

$$\sigma_{P,j,\mathbf{a}}^2 = \langle r \rangle |\mathbf{m}_j| \cdot \mathbf{p} \quad (3.27)$$

where we introduce the notation  $|\mathbf{m}_j|$  to be a vector of the absolute values of the  $j$ th row of  $M$ . The notation  $|\mathbf{m}_j|$  is distinct from the norm of a vector,  $\|\mathbf{m}_j\|$ , which returns a scalar. It is necessary here to take the absolute value of the entries in  $M$  to ensure that all variances add and do not subtract. The error in the decoded estimates is  $\mathbf{e}_{P,\hat{\mathbf{p}}} = M^{-1} \mathbf{e}_{P,\mathbf{a}}$ . The MSE of the  $k$ th reconstructed estimate is thus

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}},k} &= \sum_j^N \{M^{-1}\}_{k,j}^2 \sigma_{P,j,\mathbf{a}}^2 \\ &= \sum_j^N \{M^{-1}\}_{k,j}^2 \langle r \rangle |\mathbf{m}_j| \cdot \mathbf{p}, \end{aligned} \quad (3.28)$$

and the average MSE is

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}}} &= \frac{1}{N} \sum_k^N \epsilon_{P,\hat{\mathbf{p}},k}, \\ &= \frac{1}{N} \sum_k^N \sum_j^N \{M^{-1}\}_{k,j}^2 \langle r \rangle |\mathbf{m}_j| \cdot \mathbf{p}, \\ &= \frac{1}{N} \sum_j^N \langle r \rangle |\mathbf{m}_j| \cdot \mathbf{p} \sum_k^N \{M^{-1}\}_{k,j}^2. \end{aligned} \quad (3.29)$$

Equation 3.29 is the most general form of the average MSE for Poisson noise. The last line in equation 3.29 shows that the average MSE is modified per-acquisition by the factor  $\sum_k^N \{M^{-1}\}_{k,j}^2$ , that is, the average MSE is a weighted average of the noise at acquisition. For the special case where the Poisson noise variances in the acquired data are approxi-

mately equal,  $\langle r \rangle |\mathbf{m}_j| \cdot \mathbf{p} \approx \langle r \rangle \mathbf{m} \cdot \mathbf{p}$  for some vector  $\mathbf{m}$ , equation 3.29 becomes

$$\epsilon_{P,\hat{\mathbf{p}}} \approx \frac{\langle r \rangle \mathbf{m} \cdot \mathbf{p}}{N} \text{Tr} [(M^T M)^{-1}], \quad (3.30)$$

which shows the similarity between equation 3.29 and equation 3.12.

The average MSE for pointwise encoding,  $M = I$ , is

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}}} &= \frac{1}{N} \sum_i \langle r \rangle p_i \sum_k \{I\}_{ki}, \\ &= \langle r \rangle \langle p \rangle. \end{aligned} \quad (3.31)$$

where  $p_i$  is the  $i$ th entry in  $\mathbf{p}$  and  $\langle p \rangle$  the mean of the  $p_i$ . As with additive noise, pointwise encoding is considered the baseline and any multiplexing method must not increase the average MSE.

The H-matrix, with the inverse of  $H$  from equation 3.3, gives

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}}} &= \frac{1}{N} \sum_j \langle r \rangle |\mathbf{h}_j| \cdot \mathbf{p} \sum_k \{H^{-1}\}_{kj}^2 \\ &= \frac{1}{N} \sum_j \langle r \rangle \sum_i p_i \sum_k \left(\frac{\pm 1}{N}\right)^2 \\ &= \langle r \rangle \langle p \rangle. \end{aligned} \quad (3.32)$$

where  $|\mathbf{h}_j|$  is a vector of the absolute values of the  $j$ th row of  $H$ . Thus H-matrix multiplexing averages the errors at reconstruction but does not reduce the Poisson error *on average*. Essentially because  $H$  is a scaling and rotation, the signal and the noise are both boosted and reduced by the same amount at acquisition and reconstruction. Hassler *et al.* (2005) obtained substantially the same result when considering Poisson noise emanating per-datum to be multiplexed under the specific case of fluorescence imaging. Here we have modelled Poisson noise variance as a function of the gross ‘weight’ measured.

For the G-matrix we have

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}}} &= \frac{1}{N-1} \sum_j \langle r \rangle |\mathbf{g}_j| \cdot \mathbf{p} \sum_k \{G^{-1}\}_{kj}^2 \\ &= \frac{4}{N^2} \frac{1}{N-1} \sum_j \langle r \rangle \sum_i p_i \sum_k \{S_{N-1}\}_{kj}^2. \end{aligned} \quad (3.33)$$

where  $|\mathbf{g}_j|$  is a vector of the absolute values of the  $j$ th row of  $G$ . Since  $S$  has approximately

$N/2$  1s and  $N/2$  0s on each row and column we have

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}}} &\approx \frac{4}{N^2} \frac{1}{N-1} \sum_j^{N-1} \langle r \rangle \sum_i^{N-1} p_i \frac{N}{2}, \quad N \text{ large,} \\ &= 2 \langle r \rangle \langle p \rangle. \end{aligned} \quad (3.34)$$

The  $G$ -matrix increases noise that is proportional to  $\mathbf{p}$  and should not be used when Poisson noise is dominant.

For the  $S$ -matrix case

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}}} &= \frac{1}{N-1} \sum_j^{N-1} \langle r \rangle |\mathbf{s}_j| \cdot \mathbf{p} \sum_k^{N-1} \{S^{-1}\}_{kj}^2, \\ &= \frac{4}{N^2} \frac{1}{N-1} \sum_j^{N-1} \langle r \rangle |\mathbf{s}_j| \cdot \mathbf{p} \sum_k^{N-1} \{G\}_{kj}^2, \end{aligned} \quad (3.35)$$

where  $|\mathbf{s}_j|$  is a vector of the absolute values of the  $j$ th row of  $S$ . We make the simplifying approximation  $p_i \approx \langle p \rangle$ , use equation 3.8 for the inverse of  $S$  and observe again that each row in  $S$  has approximately  $N/2$  1s and  $N/2$  0s to find

$$\begin{aligned} \epsilon_{P,\hat{\mathbf{p}}} &\approx \frac{4}{N^2} \frac{1}{N-1} \sum_j^{N-1} \langle r \rangle \frac{1}{2} \sum_i^{N-1} \langle p \rangle \sum_k^{N-1} (\pm 1)^2, \\ &\approx 2 \langle r \rangle \langle p \rangle, \text{ for large } N. \end{aligned} \quad (3.36)$$

The  $S$ -matrix, like the  $G$ -matrix, is not suitable for Poisson noise. The result in equation 3.36 agrees with Damaschini (1993) and Nitzsche and Riesenbergs (2003).

The result that  $S$  and  $G$  have the same effect on Poisson noise contrasts with the result for additive noise (Section 3.2.1). The reason is that  $G$  increases the Poisson noise variance more than  $S$  does at acquisition but at reconstruction  $G$  reduces the Poisson variance more than  $S$ . In contrast, additive noise is independent of the encoding, so the reduction of noise by  $G$ , being greater than  $S$ , is better for additive noise.

### 3.2.4 Multiplicative Drift, Systematic Multiplicative Bias

Allowing  $\langle r \rangle$  in section 3.2.3 to vary between acquisitions models multiplicative drift. We write  $\langle r \rangle_j$  for the  $j$ th acquisition and load each  $\langle r \rangle_j$  onto the  $j$ th row of the diagonal matrix  $R$ . The average of the  $\langle r \rangle_j$  is  $\langle r \rangle$ . Assuming the acquisition to be free of any other error

$$\mathbf{a} = R M \mathbf{p}. \quad (3.37)$$

Reconstruction of the estimate  $\hat{\mathbf{p}}$  gives

$$\hat{\mathbf{p}} = M^{-1} R M \mathbf{p}, \quad (3.38)$$

which is contaminated by a bias. The error due to the bias  $\mathbf{b}$  is computed by taking the difference (Streeter *et al.*, 2008b)

$$\begin{aligned} \mathbf{b} &= \langle r \rangle \mathbf{p} - \hat{\mathbf{p}} \\ &= \langle r \rangle \mathbf{p} - M^{-1} R M \mathbf{p}, \\ &= M^{-1} (\langle r \rangle I - R) M \mathbf{p}. \end{aligned} \quad (3.39)$$

so  $\mathbf{b}$  is proportional to the differences  $\langle r \rangle - \langle r \rangle_j$ . A simple and reliably precise way to correct for drift is to observe the  $\langle r \rangle_j$  over the acquisition and correct each  $a_j$  in  $\mathbf{a}$  before reconstruction. The average squared error due to multiplicative drift is

$$\begin{aligned} \epsilon_{\mathbf{b}, \hat{\mathbf{p}}} &= \frac{1}{N} \mathbf{b}^T \mathbf{b} \\ &= \frac{1}{N} \mathbf{p}^T M^T (\langle r \rangle I - R) (M^{-1})^T M^{-1} (\langle r \rangle I - R) M \mathbf{p} \end{aligned} \quad (3.40)$$

Prediction of the relative effect of multiplexing matrices on  $\epsilon_{\mathbf{b}, \hat{\mathbf{p}}}$  cannot be performed without knowledge of the nature of the drift represented in  $R$ . Since  $R$  is different between acquisitions further general analysis is difficult.

Multiplicative drift modifies the Poisson noise variance. Including drift, equation 3.27 becomes

$$\sigma_{P, j, \mathbf{a}}^2 = \langle r \rangle_j |\mathbf{m}_j| \cdot \mathbf{p}, \quad (3.41)$$

thus the average MSE at decoding for Poisson noise with drift is

$$\epsilon_{P, \hat{\mathbf{p}}} = \frac{1}{N} \sum_j \langle r \rangle_j |\mathbf{m}_j| \cdot \mathbf{p} \sum_k \{M^{-1}\}_{kj}^2. \quad (3.42)$$

For pointwise encoding the average MSE is

$$\epsilon_{P, \hat{\mathbf{p}}} = \frac{1}{N} \sum_i \langle r \rangle_i p_i \approx \langle r \rangle \langle p \rangle. \quad (3.43)$$

For the Hadamard matrices, examination of equations 3.32, 3.33 and 3.36 shows that the  $\langle r \rangle_j$  are averaged in each case. Thus, including drift, the average MSE for the Hadamard matrices is unchanged from the average MSE of Poisson noise without drift.

### 3.2.5 Combined Effect

Combining the error effects in multiplexing gives the general acquisition model,

$$\mathbf{a} = (R + E)M(\mathbf{p} + \mathbf{t}_1) + \mathbf{t}_2 + \mathbf{e}_a. \quad (3.44)$$

As described in section 3.2.2, correction of error due to  $\mathbf{t}_1 \neq 0$  results in noise that is not affected by the multiplexing. Thus we utilise multiplexing where the encoding and the additive bias are independent so  $\mathbf{t}_1 = 0$ . We assume that  $R$  and  $\mathbf{t}_2$  can be independently measured, that is, we have access to

$$\hat{R} = R + \hat{E}_P + \hat{E}_G, \quad (3.45)$$

and

$$\hat{\mathbf{t}}_2 = \mathbf{t}_2 + \mathbf{e}_2, \quad (3.46)$$

where all  $\hat{E}_P$  and  $\hat{E}_G$  are diagonal matrices containing the Poisson and additive error in  $\hat{R}$  and the  $E_G$ ,  $\mathbf{e}_a$  and  $\mathbf{e}_2$  come from similar but independent noise sources, thus all have the same variance,  $\sigma^2$ , as each other.

Correction of the encoded measurements using the measured estimates of  $R$  and  $\mathbf{t}_2$  gives  $\mathbf{a}_c$ , namely

$$\begin{aligned} \mathbf{a}_c &= \hat{R}^{-1}(\mathbf{a} - \hat{\mathbf{t}}_2), \\ &= (R + \hat{E}_P + \hat{E}_G)^{-1} (RM\mathbf{p} + EM\mathbf{p} + \mathbf{e}_a - \mathbf{e}_2). \end{aligned} \quad (3.47)$$

Writing the combined random error in the corrected acquisition be  $\mathbf{e}_{a,c}$  gives

$$\mathbf{a}_c = M\mathbf{p} + \mathbf{e}_{a,c}, \quad (3.48)$$

and decoding gives the estimate

$$\hat{\mathbf{p}} = \mathbf{p} + M^{-1}\mathbf{e}_{a,c}. \quad (3.49)$$

Writing  $\mathbf{e}_{a,c}$  in simple form is problematic due to the matrix inversion in equation 3.47, however we may use the equation for the random error in a ratio (Pearson, 1897) to obtain the average MSE in the reconstructed corrected estimate, that is

$$\epsilon_{\hat{\mathbf{p}}} \approx \frac{1}{\langle \hat{R} \rangle^2} \left( \sigma_a^2 + \frac{\langle \mathbf{a} \rangle^2}{\langle \hat{R} \rangle^2} \sigma_{\hat{R}}^2 - 2 \frac{\langle \mathbf{a} \rangle}{\langle \hat{R} \rangle} \sigma_{a,\hat{R}}^2 \right) \frac{1}{N} \text{Tr} [(M^T M)^{-1}], \quad (3.50)$$

where  $\sigma_a^2$  is the variance in the acquisition,  $\sigma_{\hat{R}}^2$  is the variance in the drift measurement

and  $\sigma_{\mathbf{a}, \hat{R}}^2$  is the covariance between the acquisition and the drift measurements. From the subtraction of  $\hat{\mathbf{t}}_2$  which doubles the random additive variance and approximating the Poisson variances in the  $a_j$  elements of  $\mathbf{a}$  to be equal to each other gives  $\sigma_{\mathbf{a}}^2 \approx 2\sigma^2 + \langle r \rangle |\mathbf{m}| \cdot \mathbf{p}$ . The variance in the drift measurements is  $\sigma_{\hat{R}}^2 \approx \langle r \rangle + \sigma^2$  and due to the assumption of independence in the measurements the covariance  $\sigma_{\mathbf{a}, \hat{R}}^2 = 0$ . The expected values are  $\langle \mathbf{a} \rangle \approx \langle r \rangle |\mathbf{m}| \cdot \mathbf{p}$  and  $\langle \hat{R} \rangle \approx \langle r \rangle$ . Substituting into equation 3.50 and rearranging we obtain

$$\begin{aligned} \epsilon_{\hat{\mathbf{p}}} &\approx \frac{1}{\langle r \rangle^2} \left( 2\sigma^2 + \langle r \rangle |\mathbf{m}| \cdot \mathbf{p} + \frac{(\langle r \rangle |\mathbf{m}| \cdot \mathbf{p})^2}{\langle r \rangle^2} (\langle r \rangle + \sigma^2) \right) \frac{1}{N} \text{Tr} [(M^T M)^{-1}], \\ &= \frac{1}{\langle r \rangle^2} (2\sigma^2 + \langle r \rangle |\mathbf{m}| \cdot \mathbf{p} + (|\mathbf{m}| \cdot \mathbf{p})^2 \langle r \rangle + (|\mathbf{m}| \cdot \mathbf{p})^2 \sigma^2) \frac{1}{N} \text{Tr} [(M^T M)^{-1}]. \end{aligned} \quad (3.51)$$

Equation 3.51 looks like a drastic reduction in the average MSE due to the term  $1/\langle r \rangle^2$  from the drift correction, however the mean squared signal is also reduced by the same amount so the resultant SNR is not reduced by this factor. In certain cases it may not be possible to measure  $\mathbf{a}$  and  $\hat{R}$  independently so  $\sigma_{\mathbf{a}, \hat{R}}^2$  in equation 3.50 may not necessarily be zero. Such situations must be treated ad hoc for the particular multiplexing equipment.

For the case of pointwise acquisition we recall from equation 3.31 that the average MSE  $|\mathbf{m}| \cdot \mathbf{p} = \langle \mathbf{p} \rangle$  Thus equation 3.51 reduces to

$$\begin{aligned} \epsilon_{\hat{\mathbf{p}}} &\approx \frac{1}{\langle r \rangle^2} (2\sigma^2 + \langle r \rangle \langle \mathbf{p} \rangle + \langle \mathbf{p} \rangle^2 \langle r \rangle + \langle \mathbf{p} \rangle^2 \sigma^2), \\ &= \frac{2\sigma^2}{\langle r \rangle^2} + \frac{\langle \mathbf{p} \rangle}{\langle r \rangle} + \frac{\langle \mathbf{p} \rangle^2}{\langle r \rangle} + \frac{\langle \mathbf{p} \rangle^2 \sigma^2}{\langle r \rangle^2}. \end{aligned} \quad (3.52)$$

For the case of the H-matrix,

$$\begin{aligned} \epsilon_{\hat{\mathbf{p}}} &\approx \frac{1}{\langle r \rangle^2} (2\sigma^2 + N \langle r \rangle \langle \mathbf{p} \rangle + N^2 \langle \mathbf{p} \rangle^2 \langle r \rangle + N \langle \mathbf{p} \rangle^2 \sigma^2) \frac{1}{N} \frac{1}{N^2} \text{Tr} [H^2], \\ &= \frac{2\sigma^2}{N \langle r \rangle^2} + \frac{\langle \mathbf{p} \rangle}{\langle r \rangle} + \frac{N \langle \mathbf{p} \rangle^2}{\langle r \rangle} + \frac{\langle \mathbf{p} \rangle^2 \sigma^2}{\langle r \rangle^2}. \end{aligned} \quad (3.53)$$

The  $N$  that appears in the third term is disadvantageous, however shown in chapter 4, certain instrumental considerations mitigate the factor of  $N$ .

For the case of the G-matrix,

$$\begin{aligned} \epsilon_{\hat{\mathbf{p}}} &\approx \frac{1}{\langle r \rangle^2} (2\sigma^2 + N \langle r \rangle \langle \mathbf{p} \rangle + N^2 \langle \mathbf{p} \rangle^2 \langle r \rangle + N \langle \mathbf{p} \rangle^2 \sigma^2) \frac{1}{N} \frac{4}{N^2} \text{Tr} [S^2], \\ &\approx \frac{4\sigma^2}{N \langle r \rangle^2} + \frac{2 \langle \mathbf{p} \rangle}{\langle r \rangle} + \frac{2N \langle \mathbf{p} \rangle^2}{\langle r \rangle} + \frac{2 \langle \mathbf{p} \rangle^2 \sigma^2}{\langle r \rangle^2}, \quad N \text{ large.} \end{aligned} \quad (3.54)$$

The S-matrix gives

$$\begin{aligned} \epsilon_{\mathbf{p}} &\approx \frac{1}{\langle r \rangle^2} \left( 2\sigma^2 + \frac{N}{2} \langle r \rangle \langle \mathbf{p} \rangle + \frac{N^2}{4} \langle \mathbf{p} \rangle^2 \langle r \rangle + \frac{N}{4} \langle \mathbf{p} \rangle^2 \sigma^2 \right) \frac{1}{N} \frac{4}{N^2} \text{Tr} [G^2], \\ &\approx \frac{8\sigma^2}{N \langle r \rangle^2} + \frac{2 \langle \mathbf{p} \rangle}{\langle r \rangle} + \frac{N \langle \mathbf{p} \rangle^2}{\langle r \rangle} + \frac{\langle \mathbf{p} \rangle^2 \sigma^2}{\langle r \rangle^2}, \quad N \text{ large}. \end{aligned} \quad (3.55)$$

Comparing the first term in equations 3.54 and 3.55 we see again that the G-matrix reduces the additive noise by twice the amount of the S-matrix. From the second term the Poisson noise is equally doubled with respect to the pointwise acquisition. For the third and fourth terms ('mixed terms' that result from the ratio) the S-matrix outperforms the G-matrix by a factor of two. Like with the H-matrix, the third and fourth terms are mitigated by instrumental considerations.

### 3.3 Graph Theory and Hadamard Multiplexing

Ratner *et al.* (2007) established a link between strongly regular graphs and S-matrices via the eigenvalue structure and the subsequent action on noise. They showed how in certain circumstances the adjacency matrix of a strongly regular graph, which describes the location of edges between vertices, provides a useful alternative to the Hadamard S-matrix. Ratner *et al.* (2007) however did not extend the link to the Hadamard G-matrix and H-matrix. Here we:

- Reverse the Sylvester S-matrix construction to obtain strongly regular graph based equivalents to the Sylvester G-matrix and H-matrix and
- Examine the A-optimality of the constructed matrices.

A strongly regular graph (SRG) (Ratner *et al.*, 2007), written  $\text{srg}(N, k, \alpha, \beta)$ , is a graph with  $N$  vertices, where each vertex has  $k$  adjacent vertices ( $k$  connections to other vertices). When two vertices are both connected to another third vertex then the two vertices are said to have a common adjacent vertex. Each adjacent pair of vertices in a SRG has  $\alpha$  common adjacent vertices and each non adjacent pair has  $\beta$  common adjacent vertices. The adjacency of all vertex pairs can be represented by a matrix of ones and zeros called the adjacency matrix  $W$ . The eigenvalue structure of the adjacency matrix is well known and is dependant on the parameters  $N, k, \alpha$  and  $\beta$ . Specifically for  $W = \text{srg}(N, k, \alpha, \beta = \alpha)$  the largest eigenvalue is  $\lambda_1 = k$  and the other eigenvalues all have the same magnitude as each other and are less than  $\lambda_1$ . The Hadamard S-matrix also has this

eigenvalue structure. Multiplexing via  $W$  results in an average MSE of

$$\epsilon_W = \frac{\sigma^2}{N} \left( \frac{1}{k^2} + \frac{(N-1)^2}{Nk - k^2} \right). \quad (3.56)$$

When  $k = \frac{N+1}{2}$ ,  $\epsilon_W \approx \frac{4}{N}\sigma^2$ , identical to the advantage achieved by  $S$ . Hadamard matrices have only been found for a specific set of integer values of  $N$ . In principle, there is no such restriction for graph matrices, however for any specific value of  $N$  there may not necessarily be a known  $\text{srg}(N, k, \alpha, \alpha)$  for the most optimal integer value of  $k$  for minimisation of  $\epsilon_W$ . In such a case one simply selects the most optimal matrix available.

There are two immediate benefits of  $W$  over  $S$  arising from the fact that  $k$  is variable in equation 3.56 (Ratner *et al.*, 2007). The first benefit is that the multiplex order  $N$  can be raised while  $k$  is lowered. Thus if  $N$  is increased by adding more objects to be measured, then lowering  $k$  will avoid potential saturation of the measurement device. The second benefit is if multiplicative Poisson noise is present and the contributions due to Poisson error and additive error (respectively  $\rho k$  and  $\sigma^2$ ) are known then we can write, assuming homogeneous contributions to the Poisson noise by all data being multiplexed,

$$\epsilon_{W,P} = \frac{\sigma^2 + \rho k}{N} \left( \frac{1}{k^2} + \frac{(N-1)^2}{Nk - k^2} \right). \quad (3.57)$$

Finding the optimal multiplex matrix  $W$  is a matter of finding the integer value of  $k$  that minimises  $\epsilon_{W,P}$ .

### 3.3.1 The ‘Link’ Between $W$ and $S$ , $G$ and $H$

The link between  $W$  and  $S$  is established through the respective eigenvalue structure (Ratner *et al.*, 2007). We show here that starting with  $W$ , and reversing the construction procedure to get a  $S$ -matrix from an  $H$ -matrix, that a  $G$ -like matrix  $W_G = J - 2W$  shares the properties of  $H$ . Also adding the extra row and column of ones to  $W_G$  to produce an  $H$ -like matrix,  $W_H$ , results in slightly reduced noise reduction properties than  $H$ . This counter intuitive result is evidence that while under certain conditions  $S$  and  $W$  share very similar eigenvalue structure, they do not share all the same properties. Regardless  $W$  provides a useful noise reduction.

#### The Matrix $W_G = J - 2W$

To begin, we recall that for an arbitrary encoding matrix  $M$  where  $M^T = M$ , the effect on noise variance due to multiplexing is

$$\epsilon_M = \frac{\sigma^2}{N} \text{Tr} [M^{-2}], \quad (3.58)$$

where  $\text{Tr}[\cdot]$  is the matrix trace. Also recall that the matrix trace is the sum of the eigenvalues, that the trace of the matrix inverse is the sum of the reciprocal eigenvalues, and that the trace of a sum of matrices is the sum of the traces. Thus for  $W_G = J - 2W$

$$\begin{aligned}\text{Tr}[W_G^{-2}] &= \text{Tr}[(J - 2W)^{-2}], \\ &= \text{Tr}[(4W^2 + J^2 - 2WJ - 2JW)^{-1}], \\ &= \text{Tr}[(4W^2 + (N - 4k)J)^{-1}].\end{aligned}\quad (3.59)$$

since  $J$  is a matrix of ones and  $W$  has on each row and column  $k$  ones and  $N - k$  zeros. There is a fundamental lemma of linear algebra which states that for the sum of two matrices  $A + B$ , where  $A$  has full rank and  $B$  has rank one, that (Miller, 1981)

$$(A + B)^{-1} = A^{-1} - \frac{1}{1 + \text{Tr}[BA^{-1}]}A^{-1}BA^{-1}.$$

Therefore, after some algebraic manipulation,

$$\text{Tr}[W_G^{-2}] = \frac{1}{4}\text{Tr}[W^{-2}] + \frac{N - 4k}{4(4 + (N - 4k)\text{Tr}[JW^{-2}])}\text{Tr}[W^{-2}JW^{-2}].\quad (3.60)$$

The solution to the term  $\text{Tr}[W^{-2}]$  is (Ratner *et al.*, 2007) (cf. equation 3.56)

$$\text{Tr}[W^{-2}] = \frac{1}{k^2} + \frac{(N - 1)^2}{Nk - k^2}.\quad (3.61)$$

To find the value of the other trace terms we see that since  $WJ = JW = kJ$ ,

$$\begin{aligned}\text{Tr}[W^{-2}JW^{-2}] &= \frac{1}{k^4}\text{Tr}[W^{-2}W^2JW^2W^{-2}], \\ &= \frac{1}{k^4}\text{Tr}[J] = \frac{N}{k^4}.\end{aligned}\quad (3.62)$$

Similarly we have

$$\text{Tr}[JW^{-2}] = \frac{N}{k^2}.\quad (3.63)$$

Thus equation 3.60 becomes

$$\text{Tr}[W_G^{-2}] = \frac{1}{4}\left(\frac{1}{k^2} + \frac{(N - 1)^2}{Nk - k^2} + \frac{N(N - 4k)}{k^2(4k^2 + N^2 - 4kN)}\right).\quad (3.64)$$

As stated the Hadamard S-matrix coincides with  $W$  when  $k = \frac{N+1}{2}$ . Making such a substitution into equation 3.64 and assuming  $N$  large then the second term in the brackets

clearly dominates, thus

$$\begin{aligned} \text{Tr} [W_G^{-2}] &\approx \frac{1}{4} \left( \frac{(N-1)^2}{\frac{N(N+1)}{2} - \frac{(N+1)^2}{4}} \right) \\ &\approx 1. \end{aligned} \quad (3.65)$$

The noise reduction due to  $W_G$  is then

$$\epsilon_{W,G} \approx \frac{1}{N} \sigma^2. \quad (3.66)$$

Equation 3.65 implies that the rows (and columns) in  $W_G$  form a nearly orthonormal set and thus nearly satisfy the Hadamard condition (equation 3.3). As the same is not true for  $G$ , equation 3.65 also implies that  $S$  is not exactly equivalent to  $W$ .

### The ‘Equivalent’ to $H$

We examine the properties of the H-like matrix constructed from  $W$ . We construct the H-like matrix,  $W_H$ , from  $W$  by adding a row and column of ones to the front of  $W_G$ , viz

$$W_H = \begin{bmatrix} 1 & \mathbf{o}^T \\ \mathbf{o} & J - 2W \end{bmatrix}, \quad (3.67)$$

where  $\mathbf{o}$  is a column vectors of ones. The square of  $W_H$  is

$$W_H^2 = \begin{bmatrix} N+1 & (1+N-2k)\mathbf{o}^T \\ (1+N-2k)\mathbf{o} & 4W^2 + (1+N-4k)J \end{bmatrix}. \quad (3.68)$$

To simplify, we make the substitutions  $V = 2W$ ,  $\alpha_H = 1 + N - 2k$  and  $\beta_H = 1 + N - 4k$ . The inverse of  $W_H^2$  is obtained by blockwise inversion (Henderson and Searle, 1981)<sup>2</sup> as

$$W_H^{-2} = \begin{bmatrix} \frac{1}{N+1} + \frac{1}{(N+1)^2} \alpha_H^2 \mathbf{o}^T \left( V^2 + \left( \beta_H - \frac{\alpha_H^2}{N+1} \right) J \right)^{-1} \mathbf{o} & -\frac{\alpha_H}{N+1} \mathbf{o}^T \left( V^2 + \left( \beta_H - \frac{\alpha_H^2}{N+1} \right) J \right)^{-1} \\ -\frac{1}{N+1} \alpha_H \left( V^2 + \left( \beta_H - \frac{\alpha_H^2}{N+1} \right) J \right)^{-1} \mathbf{o} & \left( V^2 + \left( \beta_H - \frac{\alpha_H^2}{N+1} \right) J \right)^{-1} \end{bmatrix}. \quad (3.69)$$

---

<sup>2</sup>the Banachiewicz equation

The trace of  $W_H^{-2}$  is then

$$\begin{aligned} \text{Tr} [W_H^{-2}] = \frac{1}{N+1} + \frac{1}{(N+1)^2} \alpha_H^2 \mathbf{o}^T \left( V^2 + \left( \beta_H - \frac{\alpha_H^2}{N+1} \right) J \right)^{-1} \mathbf{o} + \\ \text{Tr} \left[ \left( V^2 + \left( \beta_H - \frac{\alpha_H^2}{N+1} \right) J \right)^{-1} \right]. \end{aligned} \quad (3.70)$$

We make the substitution  $\gamma = \beta_H - \frac{\alpha_H^2}{N+1}$  and using the matrix inverse lemma find that

$$\begin{aligned} (V^2 + \gamma J)^{-1} &= \frac{1}{4} \left( W^{-2} - \frac{k^2 \gamma}{1 + \gamma N} W^{-2} J W^{-2} \right) \\ &= \frac{1}{4} \left( W^{-2} - \frac{\gamma}{k^2(1 + \gamma N)} J \right), \end{aligned}$$

which substituted back into equation 3.70 and solving gives

$$\begin{aligned} \text{Tr} [W_H^{-2}] &= \frac{1}{N+1} + \\ \frac{1}{4} \left( \frac{1}{k^2} + \frac{(N-1)^2}{Nk - k^2} + \frac{\alpha_H^2}{(N+1)^2} \left( \frac{N}{k^2} - \frac{\alpha_H N^2}{k^2(4k^2 + \gamma N)} \right) - \frac{\gamma N}{4k^2 + \gamma N} \frac{1}{k^2} \right). \end{aligned} \quad (3.71)$$

Making the substitution  $k = \frac{N-1}{2}$ , we observe that  $\alpha_H$  and  $\gamma$  are of the order  $J(N)$ . Thus, again, the second term in the brackets dominates so

$$\text{Tr} [W_H^{-2}] \approx 1. \quad (3.72)$$

There are terms in equation 3.71 that are of order  $J(N^{-1})$  whereas there are no such terms in equation 3.65, thus the approximations made for the trace of  $W_G$  are more precise than for  $W_H$ . Typically one should expect  $\text{Tr} [W_H^{-2}]$  to be slightly greater than  $\text{Tr} [W_G^{-2}]$ .

### A Numerical Example

The  $\text{srg}(16, 6, 2, 2)$ <sup>3</sup> is an example of a SRG that nearly fits the assumptions  $k = \frac{N+1}{2}$  and  $\alpha = \beta$ . The assumption  $\alpha = \beta$  ensures the desired eigenvalue structure (Ratner *et al.*,

<sup>3</sup><http://www.maths.gla.ac.uk/~es/srgraphs.html>, date accessed 8/5/2008.

2007) and is the more important of the two. The matrix in question is

$$W = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}. \quad (3.73)$$

The largest eigenvalue of  $W^2$  is 36 and the rest are 4, which, denoting the eigenvalues as  $\lambda_i$ , give

$$\text{Tr} [W^{-2}] = \sum_i \frac{1}{\lambda_i^2} = 3.7, \quad (3.74)$$

where  $\dot{x}$  denotes recurrence. Thus  $W$  provides slightly better noise reduction than a S-matrix of order 16, if indeed one existed. The eigenvalues of  $W_G^2$  are all 16, thus

$$\text{Tr} [W_G^{-2}] = 1, \quad (3.75)$$

as predicted. The matrix  $W_H$  is of size  $17 \times 17$ . The largest eigenvalue of  $W_H^2$  is 38.651, the smallest is 10.349 and the other fifteen eigenvalues are 16. This gives

$$\text{Tr} [W_H^{-2}] = 1.06, \quad (3.76)$$

slightly higher than for  $W_G$  but still corresponding to an excellent multiplex advantage.

### 3.4 Compressed Sensing

Compressed sensing (CS) (Donoho, 2006; Tsaig and Donoho, 2006) is a modern multiplexing method for signal acquisition that requires fewer measurements than traditional

techniques. CS is based on the assumption that a signal can be represented in some transformation by a sparse vector. If one takes an orthogonal transformation matrix  $\Psi$  and computes the coefficient vector as the solution to  $\mathbf{p} = \Psi\theta$ , then  $\theta$  is sparse if many of the coefficients are zero. Frequently a transformation is not sparse but is still compressible, where most of the essential information in  $\theta$  to reconstruct  $\mathbf{p}$  is contained in a small number of coefficients. These essential coefficients are much larger in magnitude than the non-essential coefficients. The magnitude of the coefficients of a compressible signal, when sorted into decreasing order, decay according to a power law.

Consider an acquisition matrix  $\Phi$  with randomly assigned ones and negative ones with fewer rows than columns and acquire data according to

$$\mathbf{a} = \Phi\mathbf{p}. \quad (3.77)$$

Substituting  $\mathbf{p} = \Psi\theta$  gives

$$\mathbf{a} = \Phi\Psi\theta. \quad (3.78)$$

The transforms  $\Phi$  and  $\Psi$  are required to be mutually independent, that is they cannot represent each other in a compressible manner. When this mutual independence holds and when  $\theta$  is sparse (or compressible), then  $\theta$ , hence  $\mathbf{p}$ , can be accurately reconstructed with high probability.

Let  $\hat{\mathbf{p}}$  be the reconstruction of  $\mathbf{p}$  via CS. Supposing we know the location of the  $s$  largest entries in  $\mathbf{p}$ , form  $\mathbf{p}_s$  by setting all but the  $s$  largest entries to zero. Assuming data are acquired with noise  $\mathbf{e}$  where  $\|\mathbf{e}\|_2 \leq \kappa$  and that  $\delta_{2s} < \sqrt{2} - 1$  then (Candès, 2008)

$$\|\hat{\mathbf{p}} - \mathbf{p}\|_2 \leq C_0 s^{-1/2} \|\mathbf{p} - \mathbf{p}_s\|_1 + C_1 \kappa \quad (3.79)$$

where  $C_0$  and  $C_1$  are constants that depend on  $\delta_{2s}$ . From the proofs by Candès (2008), let  $\alpha_{cs} = 2\sqrt{1 + \delta_{2s}}/(1 - \delta_{2s})$  and  $\rho_{cs} = \sqrt{2}\delta_{2s}/(1 - \delta_{2s})$  then

$$\begin{aligned} C_0 &= 2 \frac{1 + \rho_{cs}}{1 - \rho_{cs}}, \\ C_1 &= 2 \frac{\alpha_{cs}}{1 - \rho_{cs}}. \end{aligned} \quad (3.80)$$

So the error in the reconstruction is bounded, even in the presence of noise.

For a signal of length  $N$ , at least  $\log N$  measurements are required for reconstruction (Donoho, 2006; Haupt and Nowak, 2006). It has also been shown empirically that if  $k$  coefficients in  $\theta$  are necessary for  $\Psi$  to represent  $\mathbf{p}$ , then about  $n = 4k$  measurements are necessary to estimate  $\theta$  to reconstruct  $\mathbf{p}$  via CS (Tsaig and Donoho, 2006).

### 3.4.1 CS and the Effect of Random Noise

Equation 3.79 explains that CS is stable in terms of accurate reconstruction. It also explains that the error contribution due to random noise in CS reconstruction increases as the signal becomes less sparse ( $s$  increases). It does not explain how the precision in CS reconstruction is affected by the number of measurements. Intuitively more measurements mean better SNR. Here we show by inference that in CS the converse is true for a random acquisition system and random additive noise, that more measurements reduces precision.

Consider the acquisition of CS data with random additive noise  $\mathbf{e}$

$$\begin{aligned}\mathbf{a} &= \Phi\mathbf{p} + \mathbf{e} \\ &= \Phi\Psi\theta + \mathbf{e}.\end{aligned}\tag{3.81}$$

Now assume that we can accurately invert the linear system,  $\Phi\Psi$ , by some linear algorithm  $\mathcal{I}[\Phi\Psi](\cdot)$ , where, for any vector  $\mathbf{x}$  satisfying the conditions of the algorithm

$$\mathcal{I}[\Phi\Psi](\Phi\Psi\mathbf{x}) = \mathbf{x},\tag{3.82}$$

thus,

$$\begin{aligned}\hat{\theta} &= \mathcal{I}[\Phi\Psi](\mathbf{a}), \\ &= \theta + \mathcal{I}[\Phi\Psi](\mathbf{e}),\end{aligned}\tag{3.83}$$

from which

$$\begin{aligned}\hat{\mathbf{p}} &= \Psi\hat{\theta} \\ &= \mathbf{p} + \Psi\mathcal{I}[\Phi\Psi](\mathbf{e}).\end{aligned}\tag{3.84}$$

Since  $\Psi$  is well conditioned and the inversion,  $\mathcal{I}$ , is linear,  $\mathcal{I}[\Phi\Psi] = \mathcal{I}[\Psi]\mathcal{I}[\Phi] = \Psi^{-1}\mathcal{I}[\Phi]$  and the random error in reconstruction is

$$\mathbf{e}_{\hat{\mathbf{p}}} = \Psi\Psi^{-1}\mathcal{I}[\Phi](\mathbf{e}) = \mathcal{I}[\Phi](\mathbf{e}),\tag{3.85}$$

thus the effect on the error in reconstruction depends only on the inversion of  $\Phi$ .

We examine the A-optimality of CS with a random matrix via

$$\epsilon = \frac{\sigma^2}{n} \text{Tr} \left[ (\Phi^T \Phi)^{-1} \right],\tag{3.86}$$

where  $n \leq N$  is the number of acquisitions taken. The inversion before the trace operation is not directly possible. The matrix  $\Phi^T \Phi$  has  $N$  eigenvalues,  $N - n$  of which are zero

which explains why  $\Phi^T \Phi$  is not invertible by traditional means. CS sidesteps the issue of invertibility, performing reconstruction with the  $n$  available pieces of information. Thus we infer  $\epsilon$  by replacing the trace with the sum of inverse of the  $n$  nonzero eigenvalues

$$\epsilon = \frac{\sigma^2}{n} \sum_i^n \frac{1}{\lambda_i^2}. \quad (3.87)$$

Eigenvalues are pairwise coupled (Ratner *et al.*, 2007). If one eigenvalue is made smaller then some other eigenvalue gets larger. The greater the spread in the eigenvalues the larger the trace and the less A-optimal CS is. Silverstein (1985) showed that the smallest and largest eigenvalues of a large Wishart matrix,  $1/N \Phi^T \Phi$ , where the entries in  $\Phi$  are independent and infinitely distributed Gaussian with variance 1, are respectively  $(1 - \sqrt{y})^2$  and  $(1 + \sqrt{y})^2$  where  $n/N \rightarrow y \in (0, 1)$  as  $n$  gets large. From Silverstein (1985) the eigenvalues of  $\Phi^T \Phi$  must be non-randomly distributed around  $N$  with finite positive support that increases with  $n$ . Thus we can infer that precision due to the A-optimality of CS decreases with more measurements. A certain number of measurements are required before reconstruction becomes accurate so there is a trade off between accuracy and precision in CS. When  $n$  is too low the error due to incorrect reconstruction dominates the error processes and when  $n$  is too high then random noise dominates.

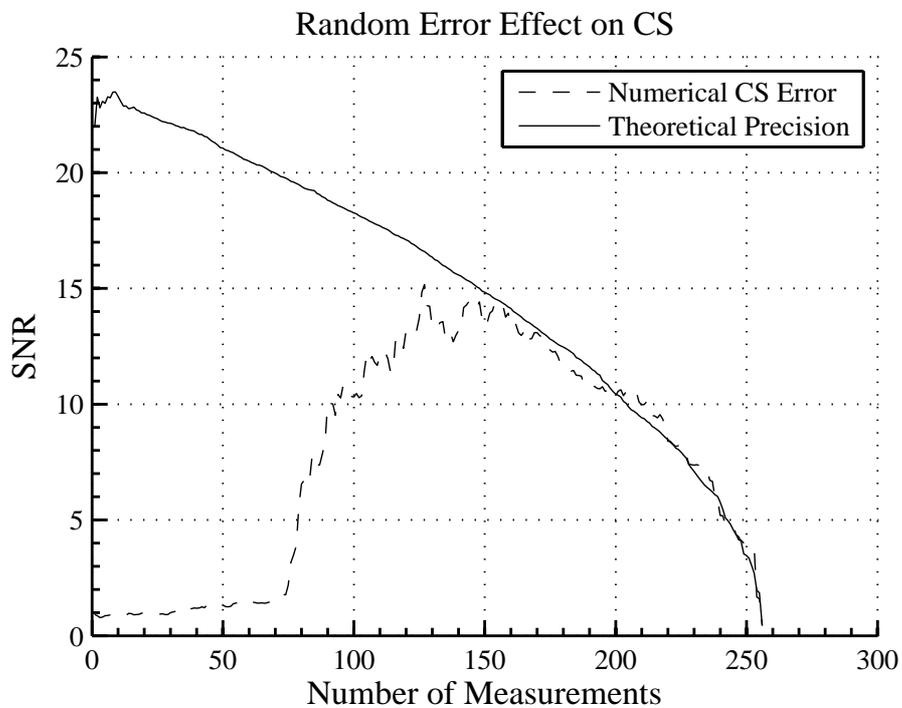


Figure 3.1: SNR of CS for a numerical example and the SNR due to the theoretically inferred precision. As the number of measurements increases the random error dominates the SNR.

We perform a numerical demonstration of the error effects in CS.

- A  $256 \times 256$  random acquisition matrix was generated with Gaussian random entries of variance one.
- A  $N = 256$  positive random signal was generated with  $k = 20$  nonzero entries.
- The encoding of the signal was contaminated with random additive Gaussian noise with variance 0.1.
- CS reconstruction was performed via basis pursuit for 1 to 256 acquisitions and for each reconstruction the SNR computed as the ratio of the root-mean-square (RMS) of the true signal over the RMS of the difference between the true signal and the reconstruction.
- The SNR due to theoretical precision was computed as the ratio of the RMS of the true signal to the square root of equation 3.87

Figure 3.1 shows the numerical CS SNR and the theoretical CS SNR due to A-optimality. The numerical SNR starts off small due to inaccurate reconstruction and then jumps quickly near  $n = 4k = 80$  measurements in accordance with Tsaig and Donoho (2006). The numerical SNR meets the theoretical SNR at 150 measurements. Above 150 measurements the random noise dominates the error and the numerical SNR tracks closely with the theoretical SNR.

### 3.5 The Effect of Error in the Reference on the Wilcoxon

Classification is the process of assigning data to one of two or more groups. For example in an automated detection system, a classifier determines if each measurement is an instance of the target for detection or not, and labels each instance as positive or negative. In performing classification tasks one must assess the power of the classifier to discriminate between groups of interest. Assessing the power of a classifier is essentially assessing how often the classifier is incorrect, however when performing the assessment one must know beforehand what group each member belongs too. The method for assigning reference groups is called the reference method and is assumed to be correct, nevertheless the reference method is another classification and cannot always be assumed to be free of error.

The area under the curve (AUC) of a receiver operator characteristic (ROC) curve and the Wilcoxon statistic are two equivalent methods for assessing the power of a classifier. When there is error in the reference then calculation of the Wilcoxon, or AUC, will be biased.

### 3.5.1 Alternative Derivation that the Wilcoxon is Equivalent to the AUC

We derive a direct link between the AUC and the Wilcoxon. The result provides a useful equation for examining the effect of error in the reference on assessing classification power. Assume some data falls into two groups, group A and group B. Group A is the ‘positive’ group and group B is the ‘negative’ group. If a classifier labels a member of group A as positive then it is a true positive, if the same classifier labels a member of group B positive then it is a false positive.

We begin with the definition of the AUC given by Barrett *et al.* (1998)

$$\text{AUC} = \int_0^1 \text{TPF}(t) d[\text{FPF}(t)]. \quad (3.88)$$

where TPF is the true positive fraction (see section 2.6.3), FPF is the false positive fraction and  $t$  is the threshold parameter that is varied to generate the ROC curve. Equation 3.88 views the TPF as a function of FPF. Taking the inverse function, i.e. the FPF as a function of the TPF, an equivalent definition of the AUC is the area between the line  $\text{FPF} = 1$  and the ROC curve

$$\text{AUC} = \int_0^1 [1 - \text{FPF}(t)] d[\text{TPF}(t)]. \quad (3.89)$$

Like Barrett *et al.* (1998) we change the variable of integration to obtain

$$\text{AUC} = \int_{-\infty}^{\infty} dt [1 - \text{FPF}(t)] \frac{d}{dt} [\text{TPF}(t)]. \quad (3.90)$$

The integration is from  $\infty$  to  $-\infty$  because TPF and FPF both go to 0 as  $t$  goes to  $\infty$  and to 1 as  $t$  goes to  $-\infty$ . The TPF is the probability that a member of A is at least  $t$ , or  $\text{TPF} = P_A(x \geq t)$ , thus

$$\frac{d}{dt} [\text{TPF}(t)] = f_A(x = t), \quad (3.91)$$

where  $f_A(x = t)$  is the frequency that  $x = t$  in group A. Similarly  $\text{FPF}(t) = P_B(x \geq t)$ , so

$$1 - \text{FPF}(t) = 1 - P_B(x \geq t) = P_B(x < t). \quad (3.92)$$

Substituting equations 3.91 and 3.92 into equation 3.90 gives

$$\text{AUC} = \int_{-\infty}^{\infty} dt P_B(x < t) f_A(t). \quad (3.93)$$

In the discrete case the members of A and B are grouped into  $N$  bins with even bin width

$L/N$ , which gives

$$\text{AUC} = \frac{L}{N} \sum_i^N P_B(x < t_i) f_A(t_i), \quad (3.94)$$

where  $L$  is the extent in the domain of the combined nonzero regions of  $f_A$  and  $f_B$  and

$$P_B(x < t_i) = \frac{L}{N} \sum_j^{i-1} f_B(t_j). \quad (3.95)$$

Now assume the bin width is made small enough so that a maximum of only one distinct value of the measurements in A and B can be present in each bin. Each  $P_B(x < t_i) f_A(t_i)$  is the normalised count of times that the members in A in bin  $i$  are larger than the members in B in bins  $j \in 1, \dots, i-1$ . Thus equation 3.94 is the normalised sum of the count of times that the members in A are larger than the members in B, which is the Wilcoxon.

### 3.5.2 The Effect of Error in the Reference

Assume group A has  $A$  members and group B has  $B$  members, with distribution of members arbitrary but clustered about the respective means. Furthermore assume that the values of the majority of members in group A are greater than the majority of members in group B, but the separation between the groups is not necessarily perfect, i.e.  $0.5 < P(A > B) \leq 1$ .

Error in the reference assigns members to the incorrect group. Let  $B^*$  be the number of members actually from group B but are missassigned to group A and  $A^*$  the number of members from group A missassigned to group B. The frequency distribution of the members erroneously transferred from group A is  $f_A^*$ , and likewise for group B,  $f_B^*$ . The unnormalised frequency distribution of group A with error is  $Af_A - A^*f_A^* + B^*f_B^*$  and the probability of group B with error is  $BP_B + A^*P_A^* - B^*P_B^*$ . The Wilcoxon where there is error in the reference is (c.f equation 3.94)

$$W_e = \frac{L}{N} \sum \frac{[BP_B + A^*P_A^* - B^*P_B^*] (x < t) [Af_A - A^*f_A^* + B^*f_B^*] (t)}{(B + A^* - B^*)(A - A^* + B^*)}. \quad (3.96)$$

The exact effect of error on the Wilcoxon depends on the distributions of member transference. We proceed with the simplest case that every member in both groups has equal chance of missassignment. Thus  $f_A^*$  and  $P_B^*$  are scaled versions of the distributions without error, namely  $f_A$  and  $P_B$ . Letting the Wilcoxon without error be  $W$  and recalling that the Wilcoxon of a group of data with itself is 0.5, gives

$$W_e = \frac{W(BA - BA^* - B^*A) + 0.5(B^*B + AA^* - (B^*)^2 - (A^*)^2 + 2A^*B^*)}{(B + A^* - B^*)(A - A^* + B^*)}. \quad (3.97)$$

An interesting feature is that due to the assumption that the errors in the reference occur with uniform probability, when  $A = B$ ,  $A^* = B^*$  and equation 3.97 becomes

$$W_e = W \left( 1 - 2 \frac{B^*}{B} \right) + \frac{B^*}{B}. \quad (3.98)$$

Substituting  $k = 2 \frac{B^*}{B}$  where  $k$  is a continuous variable between 0 and 1 gives

$$W_e = W (1 - k) + \frac{k}{2}, \quad (3.99)$$

which is a linear interpolation between the points  $(k, W(k)) = (0, W)$  and  $(k, W(k)) = (1, 0.5)$ , i.e. a straight line between the Wilcoxon with no error in the reference and random group assignment.

The importance of this theoretical result is that given the Wilcoxon and an estimate of the amount of error in the reference  $B^*$  and  $A^*$ , then the Wilcoxon without error is estimated by projection back along equation 3.97 to the point  $A^* = B^* = 0$ . Alternatively rearranging equation 3.97 gives

$$W = \frac{W_e (B + A^* - B^*) (A - A^* + B^*) - 0.5 (B^* B + A A^* - (B^*)^2 - (A^*)^2 + 2 A^* B^*)}{(B A - B A^* - B^* A)}. \quad (3.100)$$

For the straight line that occurs when  $A = B$ ,

$$W = \frac{W_e - B^*/B}{1 - 2B^*/B}. \quad (3.101)$$

The uncertainty of the estimate of the true  $W(A^* = B^* = 0)$  is determined by the uncertainty in the computed  $W_e$  and the uncertainty in the estimate of  $A^*$  and  $B^*$ . Let  $\theta$  be the true AUC, the standard error in  $W$  is (Hanley and McNeil, 1982)

$$SE(W) = \frac{\sqrt{\theta(1 - \theta) + (A - 1)(Q_1 - \theta^2) + (A - 1)(Q_2 - \theta^2)}}{AB}, \quad (3.102)$$

where " $W$  can be thought of as an estimate of  $\theta$ ..." (Hanley and McNeil, 1982),  $Q_1$  is the probability that two randomly chosen members of group A will have higher value than any random member of group B and  $Q_2$  is the probability that one randomly chosen member of group A will be higher than any two randomly chosen members of group B:

$$\begin{aligned} Q_1 &= \frac{\theta}{2 - \theta}, \\ Q_2 &= \frac{2\theta^2}{1 + \theta}. \end{aligned} \quad (3.103)$$

For  $W_e$  we might make the appropriate substitutions for  $A$  and  $B$  in equation 3.102 to include  $A^*$  and  $B^*$ , however in practise one would use equation 3.102 with the awareness that the values of  $A$  and  $B$  used include the error. Computing the uncertainty of  $W$  from the uncertainty in  $W_e$  and  $B^*$  is problematic because we do not have accurate knowledge of  $B$ , rather, we have  $B - B^* + A^*$ . Instead graphical means are employed where curves are computed for the upper and lower limits of the estimate of  $W_e$  and  $B^*$ , with the point  $W_e(B^*/B = 0.5, A^*/A = 0.5) = 0.5$  taken as definite, i.e the assumption that when half the data points are transferred into the other group the Wilcoxon is 0.5.

A very important point is that the error in the reference may not occur completely at random so the distributions  $f_A^*$  and  $f_B^*$  may not be scaled versions of  $f_A$  and  $f_B$ . Since the reference method used for classifier assessment is itself a classifier then errors in the reference are more likely to occur where the groups overlap. The salient issue is if the reference method is independent to the new method. Considering each group individually, if the reference method and the new method are uncorrelated then the reference errors occur with equal probability in the new method. If they are correlated then the reference errors will not appear randomly in the new method, rather they will be biased towards either the high or low end of the distribution. If the reference method is biased towards correct classification of either group then missassignment of the members of the other group will occur more frequently. Furthermore if data cleaning is performed per group, then data points outside of the overlap region and transferred from one group to the other may appear as outliers and be removed, changing the distribution of transferred data. Maximum likelihood estimation with assumptions about the underlying distributions to estimate  $f_A^*$  and  $f_B^*$  is conceivable but beyond the scope of this thesis.

### 3.5.3 Numerical Simulation

We demonstrate the effect of reference error on the Wilcoxon by numerical simulation and match the results with the theoretical prediction. Two groups of data are generated, the first set is designated group A and has  $A = 3000$  members. The second group is group B and has number of members  $B = k_B A$ , where  $k_B$  can be any real number greater than zero. Both groups are normally distributed with variance 1. Group B has mean 0 where Group A has mean 1.5.

The following procedure was performed.

1. For each of  $k_B$  in the set  $\{0.3, 1, 10\}$  let  $B = k_B A$  and generate groups A and B.
2. For  $p$  from 1 to 50%, randomly select and transfer  $p$  members from group A to group B,  $p$  members from group B to group A, and compute the Wilcoxon.

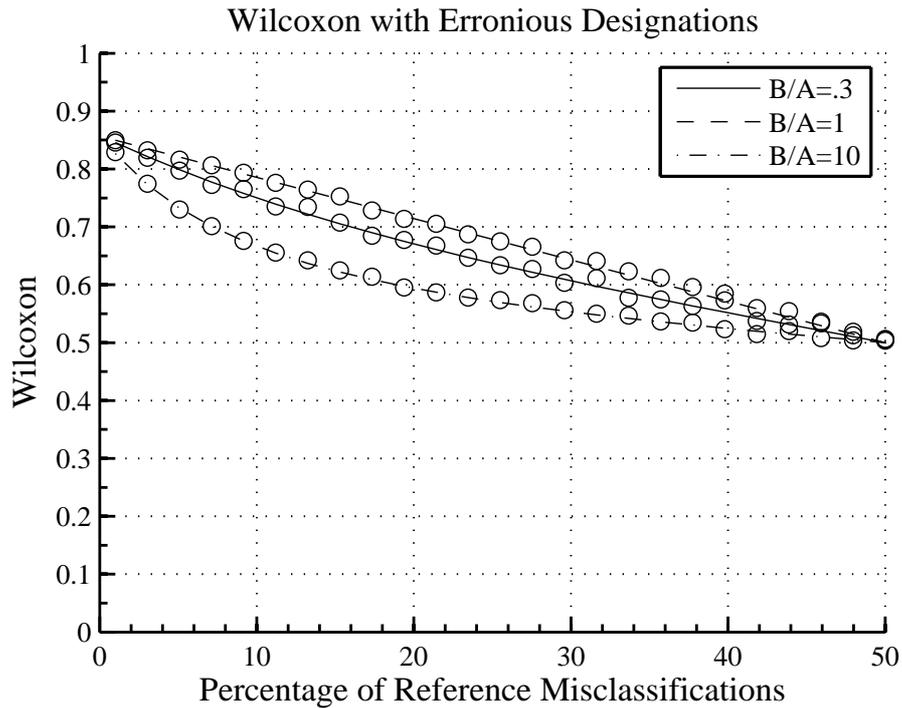


Figure 3.2: Numerical computation and theoretical prediction of the Wilcoxon with varying amount of error in the reference.

Figure 3.2 shows the result of the numerical simulation accompanied with the theoretical prediction for each  $k_B = B/A$ . The numerical data points are presented as circles where the different theoretical lines are identified by their respective line types. The true Wilcoxon with no reference error was 0.86. The theoretical prediction matches the numerical simulation well, where it is clear in figure 3.2 which theoretical line matches which set of numerical data points.



# Chapter 4

## Hardware

This chapter describes the design, theory and use of the physical system. The hardware consists of an optical light projection system to encode light patterns on the sample and a visible/near infrared spectrometer to collect reflected light. The projector uses a Texas Instruments Digital Micromirror Array (DMA) as the optical engine. The DMA was chosen because it is fully controllable at individual mirror level and for the broad spectral characteristics of the mirrors.

Before modern light modulation devices such as DMAs, light modulation was traditionally implemented via mechanical multiple aperture gratings and grids. Cyclic matrices were traditionally very popular in optical multiplexing because a physically self supporting mechanical grid could be constructed according to the first row, where self supporting means that the grid holds itself together. We explain the principle in more detail. For a cyclic S-matrix generated from a cyclic sequence of length  $2^m - 1 = ab$ , the sequence is wrapped into a matrix of dimensions  $a \times b$ . Each row of the new matrix is a length  $b$  fragment of generating sequence. The code fragment on each row of the  $a \times b$  wrapped matrix is continued to the right to generate an  $a \times ab$  matrix. Each row of this new matrix is the same row above it but shifted cyclically  $b$  times. The first  $a \times b$  ‘window’ of the new matrix is the wrapped first row of the S-matrix, the second  $a \times b$  ‘window’ the wrapped second row and so on. Careful selection of the S-matrix and the dimensions  $a$  and  $b$  ensures that the wrapped cyclic matrix is self supporting, namely that every ‘1’ entry is vertically, horizontally and/or diagonally adjacent to another ‘1’, likewise every ‘0’ is adjacent to another ‘0.’ A physical self supporting grid of holes and occlusions to represent the 1’s and 0’s can therefore be made to represent a cyclic S-matrix.

Modern light modulation devices completely remove the need for self supporting encoding schemes. Figure 4.1 shows an image of a DMA chip (with a picture drawn on the mirror region) and a close up drawing of nine mirrors on the chip. Each mirror in a DMA can be individually addressed and programmed to deflect into the nominally on or off state (respectively towards the optical path or away into free space or a light dump).

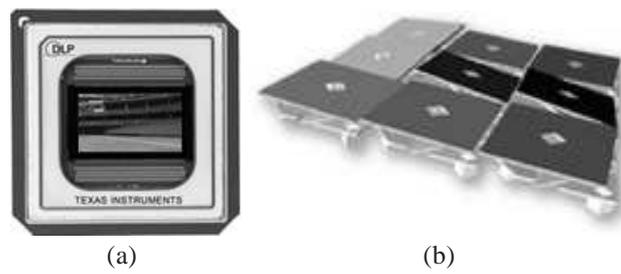


Figure 4.1: (a) A DMA chip and (b) a close up drawing of the mirrors. (Source: <http://www.dlp.com/tech/what.aspx>, date accessed 16/02/2009, ©Copyright 2009 Texas Instruments Incorporated. Used by permission.)

Any binary encoding pattern of order equal to or less than the number of mirrors in the DMA can conceivably be used (barring aliasing). A typical DMA might have a grid of  $800 \times 600$ ,  $1024 \times 768$  or even more mirrors. The degree of flexibility permitted allows the use of non-cyclic Hadamard matrices (Streeter *et al.*, 2009) and the random patterns required for compressed sensing (Streeter *et al.*, 2008a). Furthermore, unlike traditional mechanical methods, the same optical system can implement Hadamard multiplexing or compressed sensing of any order without replacement of any parts.

## 4.1 Optics and Operational Theory

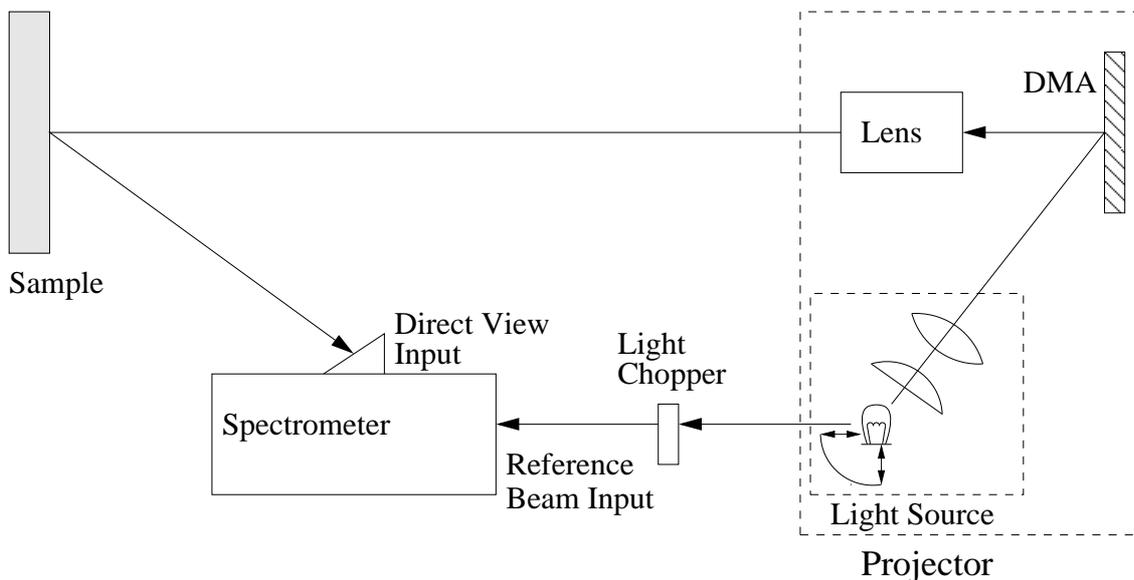
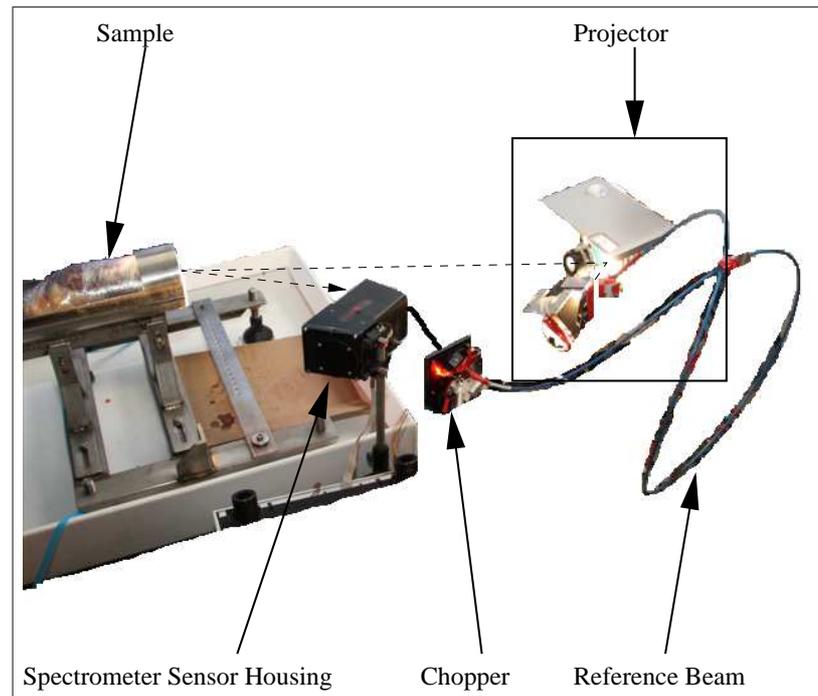
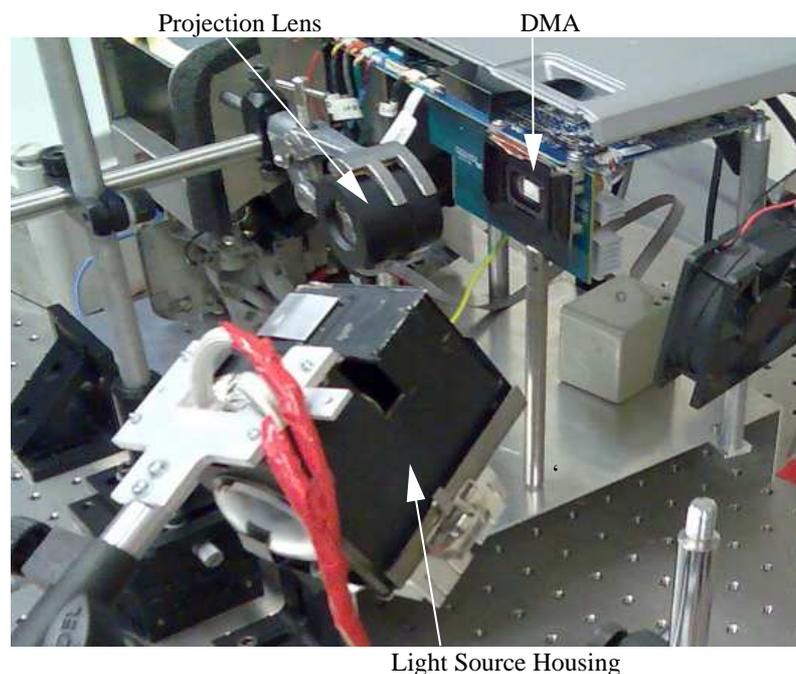


Figure 4.2: Diagram of the optical setup. Arrows indicate the light path. The spectrometer measures light from the sample and light piped directly from the source. The light chopper rate is controlled by the spectrometer.

Figure 4.2 shows the optical arrangement. For the illumination system a custom



(a)



(b)

Figure 4.3: a) The optical arrangement, the dashed arrow indicates the sample light path. The DMA based custom light projector is identified at the top right and is comprised of the digital projector, with optics stripped down to the DMA and replaced with slide projector optics. A sample (a piece of meat) is seen on the left. The spectrometer sensor housing is on a pole to the right of the sample in the image. The blue optical fibre is the reference beam, conducting light directly from the source to the light chopper. b) A close up detail of the custom projector without the reference beam fibre. The large black box with the red taped wires is the light bulb housing that contains the backreflector, condenser and shaper optics. The light source is pointed at the mirrors on the DMA and the projection lens is held in place in front the the DMA.

broadband digital micromirror array (DMA) projector was built (Streeter *et al.*, 2007, 2008b, 2009). The projector illuminates the sample with multiplexing patterns, encoding the spatial information of the sample. To build the custom projector, the optics of a commercially available digital data projector (PD100D, ACER, Taiwan) were stripped back to provide full access to the DMA, and replaced with a light source suitable for Vis/NIR spectroscopy and an appropriate projection lens. The light source, DMA and projection lens together form the projector. The particular DMA has a grid of  $800 \times 600$  mirrors. The light source is a 250 W tungsten halogen bulb (Osram, Germany) with a back reflector and condenser and shaper lenses from a Hanimex Syllabus 2000 slide projector (Hanimex Australia, now part of Fujifilm Australia<sup>1</sup>). The Hanimex light source illuminates the DMA. Mirrors on the DMA that are set to the on state reflect light through a 50 mm projection lens taken from a Leitz Pradovit 153DU slide projector (Ernst Leitz AG, a former German corporation), which images the mirror pattern on to the sample. Light incident on DMA mirrors in the off state deflect into free space, landing on the ceiling of the room a good ten metres from the DMA and out of view of the spectrometer. The custom projector illuminates the sample with a ‘view’ window at the sample of approximately  $100 \text{ mm} \times 130 \text{ mm}$ . S-matrix encoding uses the full 130 mm width whereas H-matrix encoding utilises the central 100 mm width of the view window.

A diode array spectrometer (100 series, KES Analysis Inc., NY, USA) collects and measures light reflected from the sample. The spectrometer unit analyses the band 400–1700 nm with wavelength spacing ranging from 6 nm at 400 nm to 10 nm at around 1300–1400 nm. The light from the sample is collected by a direct view input, which is simply a collection lens in front of the entrance slit of the spectrometer. The collection lens does not appear to precisely image the sample onto the entrance slit, otherwise the spectrometer would only sample light from a thin slit on the sample, rather the lens is partially out of focus with the entrance slit so that a blurred area is sampled. The blurred entrance slit image on the sample is then an oval collection region. The direct view input is standard on the KES 100 series spectrometer and collects light from the numerical aperture cone in front of it. The spectrometer is positioned as close to the sample as possible to collect a maximum amount of light, but is as far away from the sample as necessary so that light from the entire sample is collected. Images were captured and examined and the spectrometer repositioned to minimise edge effects (minimise the presence of corner pixels from which light does not reach the sensor).

The spectrometer has an inbuilt light chopper arrangement that is synchronised with the spectrometer’s data acquisition sequence. The chopper was detached from the main spectrometer body and mounted externally for full access. A gradient index lens focuses light from the halogen bulb of the Hanimex light source into a 1 mm diameter optical fibre

---

<sup>1</sup><http://hanimex.co.nz/>, date accessed 02/07/2009

that conducts the light from the light source to the chopper. The chopped reference beam is conducted to the spectrometer via a fibre optic bundle. Due to physical constraints the light to the sample is not chopped. When the chopper is closed the spectrometer receives light from the sample only, when the chopper is open the spectrometer receives light from the sample and the reference beam.

The DMA and spectrometer are computer controlled via Matlab (v2007a, The Mathworks, MA, USA). KES Inc. ActiveX libraries provide the interface between Matlab and the Spectrometer. The ACER data projector controller hardware is left intact and the DMA chip controlled via the computer VGA interface. The Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) is used to interface Matlab with the projector.

A physical masking system is not used to prevent light from outside of the sample to reach the entrance slit. Instead, assuming that unwanted light is not changing between sequent spectral acquisitions, the data acquisition and reconstruction method described below (section 4.1.2) removes unwanted light, producing a virtual masking effect. There is ample opportunity for the optical arrangement to cause unwanted spatial and spectral effects, such as attenuation at optical surfaces and uneven sampling of the sample by the spectrometer: the centre is seen as brightest, even if the sample is evenly illuminated. An internal study performed at AgResearch independent of this thesis observed the uneven collection with the centre being brightest.<sup>2</sup> These effects are multiplicative in nature and are easily corrected (see Section 4.1.5 below).

Figure 4.3 shows a) an image of the experimental setup and b) a closeup of the custom image projector. The experimental setup is the final setup used for the meat spectral imaging application in chapter 5. The spectral imaging is in process in a), so the light source is on and the DMA cannot be seen because of the bright light illuminating it. The picture of the close up of the custom projector in b) was taken early in the setup process, so a minimum number of components are seen (in particular the reference beam is absent, which was added later). The light source housing (with the red taped wires) and projection lens (held in place with a retort stand arm) are clearly seen in the picture. The DMA is the rectangular grey object to the right of the projection lens. There is a black mask object in front of the DMA that permits only the mirror region to be seen.

### 4.1.1 Data Acquisition

Before discussing multiplexing on the optical system we explain acquisition of one spectrum with an associated reference beam measurement. Every acquired measurement comprises of two parts: the spectrum via the sample,  $a$ , and the spectrum including the light source reference,  $a_r$ . The acquired sample spectrum,  $a$ , consists of: the sample spectrum  $p$

---

<sup>2</sup>Private Communication.

(an attenuation), the light source spectrum  $r$  (with associated noise  $e_P$  of variance  $r$ ), the background  $t$  which includes stray light and the spectrometers baseline and the random additive noise  $e$ . Now

$$a_\lambda = \alpha_{s,\lambda}(r_\lambda + e_{P,\lambda})p_\lambda + t_\lambda + e_\lambda, \quad (4.1)$$

where  $\lambda$  is the index of wavelength observed by any particular diode of the spectrometer and  $\alpha_{s,\lambda}$  is the attenuation of light due to the optics. As stated above, due to physical constraints with the available equipment, the light to the sample is not chopped whereas the reference beam is. Therefore the second part of the spectrum,  $a_r$ , includes, in addition to the reference beam,  $r$ , and the associated random additive noise,  $e_r$ , the sample spectrum  $a$ . Thus

$$a_{r,\lambda} = \alpha_{r,\lambda}(r_\lambda + e_{P,r,\lambda}) + \alpha_{s,\lambda}(r_\lambda + e_{P,r,\lambda})p_\lambda + t_\lambda + e_{r,\lambda}. \quad (4.2)$$

where  $\alpha_{r,\lambda}$  is the attenuation through the reference optics and the subscript  $r$  denotes reference. The time difference between the acquisition of  $a$  and  $a_r$  is assumed small enough that  $t$  does not change between the two measurements. The estimate of the reference,  $\hat{r}$ , is recovered by subtraction

$$\hat{r}_\lambda = a_{r,\lambda} - a_\lambda = \alpha_{r,\lambda}r_\lambda + (\alpha_{r,\lambda} + \alpha_{s,\lambda})e_{P,r,\lambda} - \alpha_{s,\lambda}e_{P,\lambda}p_\lambda + e_{r,\lambda} - e_\lambda. \quad (4.3)$$

With any reference beam system the optics are designed to pass as much light as possible, so  $\alpha_{r,\lambda} \gg \alpha_{s,\lambda}$  and  $\alpha_{r,\lambda}e_{P,r,\lambda} \gg e_{r,\lambda} - e_\lambda$  and we can approximate the reference estimate as

$$\hat{r}_\lambda \approx \alpha_{r,\lambda}r_\lambda + \alpha_{r,\lambda}e_{P,r,\lambda}. \quad (4.4)$$

The subtraction between  $a_r$  and  $a$  in equation 4.3 causes  $a$  and  $\hat{r}$  to have non-zero covariance. The simplification made in equation 4.4 does not mean that the covariance is approximately zero. The covariance  $\sigma_{a,\hat{r}}^2$  is the variance of the noise terms that appear in both equations 4.1 and 4.3. Let  $\sigma_0^2$  be the variance of the random additive noise and since the variance of the source term is equal to the mean  $r$ , the covariance is

$$\sigma_{a,\hat{r}}^2 = -\alpha_s r p_\lambda - \sigma_0^2. \quad (4.5)$$

### 4.1.2 Complement Encoding

Consider the Hadamard encoding matrix  $H$  with  $+1$  and  $-1$  entries. The optical system of Figure 4.2 cannot directly implement the  $-1$ 's of  $H$ . We split the encoding matrix  $H$  into two parts, each with entries of 0 and 1: a positive part  $H^+$  and negative part  $H^-$  such

$$\begin{array}{rcc}
\mathbf{H} & \begin{array}{|c|c|c|c|c|} \hline +1 & +1 & -1 & +1 & -1 \\ \hline \end{array} & = \\
\mathbf{H}^+ & \begin{array}{|c|c|c|c|c|} \hline +1 & +1 & 0 & +1 & 0 \\ \hline \end{array} & - \\
\mathbf{H}^- & \begin{array}{|c|c|c|c|c|} \hline 0 & 0 & +1 & 0 & +1 \\ \hline \end{array} & 
\end{array}$$

Figure 4.4: Illustration of splitting the Hadamard matrix into  $H^+$  and  $H^-$  components. In  $H^+$  the  $-1$ s of the original  $H$ -matrix are converted to 0's. In  $H^-$  the  $+1$ s are converted to 0s and the  $-1$ 's are converted to  $+1$ s.

that  $H = H^+ - H^-$ , where

$$\begin{aligned}
H^+ &= \frac{1 + H}{2}, \\
H^- &= \frac{1 - H}{2}.
\end{aligned}$$

The  $H^+$  and  $H^-$  encode complementary subsets of the pixels. Figure 4.4 illustrates the splitting process. Davis (1995) used occlusion and reflection to split the Hadamard matrix to perform hyperspectral imaging. The optics of our system in conjunction with the splitting of the multiplexing matrix facilitates background illumination removal (Streeter *et al.*, 2007), which was not considered by Davis. The splitting process was also proposed in optical communications for code division multiple access systems (Nguyen and Young, 1995) to boost the SNR of identifying a single user from multiple signals.

The spectrometer measures a set of bands, each labelled with a particular wavelength. We drop the subscript  $\lambda$  as the operations described are applied simultaneously and independently to each measured band. To acquire the spectra the  $j$ th row of  $H^+$  is wrapped into a two dimensional pattern, projected onto the sample and the corresponding  $j$ th entry of  $\mathbf{a}^+$  is acquired. Immediately after each row of  $H^+$  the corresponding row of  $H^-$  is similarly wrapped and projected onto the sample and the corresponding  $j$ th entry of  $\mathbf{a}^-$  acquired. The spectra are contaminated by random additive instrument noise  $\mathbf{e}$  and the additive combination of dark current, background and stray light, all represented by  $T$ . The light source in our experimental setup has been shown to be stable during the data acquisition period (Streeter *et al.* (2008b), also see Section 4.3.2), as no significant improvement in measurement repeatability was seen after correcting for light source drift in spectral regions of usable SNR in that study. Regardless we model the effect of light source drift during the data acquisition as multiplication by the diagonal matrix  $R$ . The

illumination has random fluctuation following Poisson statistics represented by the diagonal matrix of random errors  $E_P$ . For pixel responses  $0 \leq \{\mathbf{p}\}_i \leq 1$  at pixel  $i$  on the sample, the acquired spectra are then:

$$\begin{aligned}\mathbf{a}^+ &= \alpha_s(R^+ + E_P^+)H^+\mathbf{p} + T + \mathbf{e}^+, \\ \mathbf{a}^- &= \alpha_s(R^- + E_P^-)H^-\mathbf{p} + T + \mathbf{e}^-, \end{aligned}\quad (4.6)$$

The variance of the diagonal entries in  $E_P$  are

$$\text{var}(\{E_P\}_{j,j}) = \{R\}_{j,j} = \langle r \rangle_j, \text{ and } \{E\}_{i,j} = 0 \text{ for all } i \neq j. \quad (4.7)$$

where  $\text{var}(\cdot)$  denotes variance. Equation 4.7 states that the illumination over the sample has random fluctuation with variance equal to the mean intensity; characteristics typified by Poisson statistics. The superscripts on  $R$ ,  $E_P$  and  $\mathbf{e}$  highlight that random noise sources change between acquisitions.

There is a second source of intensity dependent noise known as shot noise, that occurs in the sensor diodes and is proportional to the current induced by photon interaction with the sensor. The effect of shot noise is a constant modifier, greater than one, on the right hand side of the first equation in 4.7. Furthermore this constant modifier is dependent on the width of the band (not truly a bandwidth in the sense that the diodes do not have a Gaussian sensitivity profile) of the diode in the same way the intensity noise is dependant on the bandwidth of the observation. In a spectrometer the band dependence may not be precisely identical between diodes and thus may cause a ‘wavelength’ dependent noise effect in the spectra. We do not explicitly model shot noise as the theoretical results and associated interpretations that follow are not altered in a significant way. Regardless we recognise that the shot noise effect exists.

The combined stray and background light,  $T$ , is assumed to be slowly changing and, because of the source encoding (equation 4.6),  $T$  is independent of  $H$ . As each row of the positive and negative encodings are taken in quick succession we make the approximation  $R \approx R^+ \approx R^-$ . Thus taking the difference between positive and negative encoding parts gives

$$\mathbf{a} = \mathbf{a}^+ - \mathbf{a}^- = \alpha_s R H \mathbf{p} + \mathbf{e}_{a,P} + \mathbf{e}, \quad (4.8)$$

where  $\mathbf{e} = \mathbf{e}^+ - \mathbf{e}^-$  is the total additive noise and

$$\mathbf{e}_{a,P} = \alpha_s(E_P^+H^+ - E_P^-H^-)\mathbf{p} \quad (4.9)$$

is the total Poisson photon noise. The variance of  $\mathbf{e}$  is  $\sigma^2 = 2\sigma_0^2$  due to the subtraction

operation. The value of  $\sigma^2$  is not to be identified or confused with the theoretical  $\sigma^2$  of the previous chapter. The encoding is performed according to matrices with positive entries and the  $H^+$  and  $H^-$  encode complement subsets of  $\mathbf{p}$ , thus the signal dependent noise variance for the  $j$ th acquisition is  $\sigma_{a,P,j}^2 = \alpha_s |\mathbf{h}_j| \cdot \mathbf{p}$  where  $|\mathbf{h}_j|$  is a vector consisting of the absolute values of the entries of the  $j$ th row of  $H$ . For the G-matrix (section 3.1.1) we substitute  $|\mathbf{g}_j|$  for  $|\mathbf{h}_j|$ .

### 4.1.3 Decoding and Noise in Complement Encoding

Application of the inverse transform to the acquired spectra gives

$$\begin{aligned} \hat{\mathbf{p}} &= \frac{1}{N} H \mathbf{a} \\ &= \frac{\alpha_s}{N} H R H \mathbf{p} + \frac{\alpha_s}{N} H (E_P^+ H^+ - E_P^- H^-) \mathbf{p} + \frac{1}{N} H \mathbf{e}. \end{aligned} \quad (4.10)$$

The last term in equation 4.10 is the reduced additive noise. The first term in equation 4.10 is the reconstructed pixel values but now contaminated by a multiplicative matrix factor. If the light source drifts over time then the factor corrupts the relative magnitude of the entries in  $\hat{\mathbf{p}}$  by

$$\frac{\alpha_s}{N} H R H. \quad (4.11)$$

If the light source does not drift then  $R = \langle r \rangle I$  so

$$\frac{\alpha_s}{N} H \langle r \rangle I H = \alpha_s \langle r \rangle I, \quad (4.12)$$

and the multiplicative error factor is constant over the entries in  $\hat{\mathbf{p}}$ .

The second term in equation 4.10 is the photon noise. If we were to illuminate the entire sample then the photon noise variance is  $\alpha_s \langle r \rangle_j N \langle p \rangle$  where  $\langle p \rangle$  is the mean pixel value,  $N \langle p \rangle$  is the attenuation of the entire imaged region and  $\langle r \rangle_j$  is the light intensity at the time that the  $j$ th acquisition is taken. Recall that each positive encoding pattern (row in  $H^+$ ) illuminates a subset of the pixels and the corresponding negative encoding from  $H^-$  illuminates the complement subset. Thus there exists the fraction  $0 < \delta_j < 1$  such that the photon noise variance of the  $j$ th acquisition of the positive encoding is  $(\sigma_{j,P}^+)^2 = \delta_j \alpha_s \langle r \rangle_j N \langle p \rangle$  and the negative encoding  $(\sigma_{j,P}^-)^2 = (1 - \delta_j) \alpha_s \langle r \rangle_j N \langle p \rangle$ . The photon noise variance  $\sigma_{j,a,P}^2$  of the  $j$ th acquisition in equation 4.8 is then

$$\sigma_{j,a,P}^2 = (\sigma_{j,P}^+)^2 + (\sigma_{j,P}^-)^2 = \alpha_s \langle r \rangle_j N \langle p \rangle. \quad (4.13)$$

The photon noise for the G-matrix at acquisition is approximately the same as equation 4.13 because  $G$  has no 0 entries.

H-matrix decoding averages the error values (with sign flipping due to the negative entries in  $H$ ) and reduces the random noise MSE by a factor of  $1/N$ . The MSE of the estimates due to photon noise at decoding (equation 4.10) is therefore

$$\epsilon_{\hat{\mathbf{p}},P} = \alpha_s \langle r \rangle \langle p \rangle. \quad (4.14)$$

Thus for the H-matrix the average MSE is

$$\epsilon = \frac{1}{N} \sigma^2 + \alpha_s \langle r \rangle \langle p \rangle + \frac{1}{N} \mathbf{b}^T \mathbf{b}, \quad (4.15)$$

where  $\mathbf{b}$  is the bias error due to the light drift (section 3.2.4) which depends on  $H$ , namely

$$\mathbf{b} = \alpha_s \left( \langle r \rangle I - \frac{1}{N} H R H \right) \mathbf{p}. \quad (4.16)$$

When the light source does not drift then from equation 4.12,  $\mathbf{b} = 0$ . Multiplicative bias that alters the relative magnitude of the estimates in  $\hat{\mathbf{p}}$  also results in nonzero  $\mathbf{b}$ . A constant bias is easily correctable, however a non constant bias is more difficult if not impossible to correct without auxiliary measurement of the cause.

Decoding for the G-matrix does not reduce the photon noise to the same degree as the H-matrix. The average MSE for G-matrix multiplexing at decoding is

$$\epsilon \approx \frac{2}{N-1} \sigma^2 + 2\alpha_s \langle r \rangle \langle p \rangle + \frac{1}{N-1} \mathbf{b}^T \mathbf{b}. \quad (4.17)$$

The bias  $\mathbf{b}$  in equation 4.17 is dependent on  $G$ . Thus the bias error in equation 4.17 is not the same as that of equation 4.15. For  $G$ ,  $\mathbf{b}$  is

$$\mathbf{b} = \alpha_s \left( \langle r \rangle I - \frac{2}{N} S R G \right) \mathbf{p}. \quad (4.18)$$

For pointwise encoding, it is necessary to acquire a measurement of the background illumination/baseline/stray with every pixel measurement. The continual background measurement is to mitigate any error effects due to background drift, but causes the random additive error in acquisition to be  $\sigma^2 = 2\sigma_0^2$ , the same as the complement encoding. The background measurement is subtracted from the pixel measurement, thus the average MSE for pointwise encoding is

$$\epsilon \approx \sigma^2 + \alpha_s \langle r \rangle \langle p \rangle + \frac{1}{N} \mathbf{b}^T \mathbf{b}. \quad (4.19)$$

Again the bias  $\mathbf{b}$  is dependent on the acquisition matrix, in this case the matrix identity

$$\mathbf{b} = \alpha_s (\langle r \rangle I - R) \mathbf{p}. \quad (4.20)$$

In the above the noise was averaged to produce a single noise characteristic, namely the A-optimality (average optimality, see section 3.2.1). Considering the pointwise photon noise MSE per pixel (which is also per acquisition for pointwise acquisition) we have  $\sigma_{j,P}^2 = \alpha_s \langle r \rangle_j \langle p \rangle_j$ . In multiplexing ( $H$  or  $G$  for example) the noise per-pixel after reconstruction is the average of the noise at acquisition. This is a subtle but important point, because the Poisson noise is increased for lower than average  $p_i$  (Hassler *et al.*, 2005). If the image has a high dynamic range then dark pixels have Poisson noise increased by multiplexing. Likewise bright pixels have Poisson noise decreased. Whether the increase of Poisson noise in dark pixels is important is application dependent.

#### 4.1.4 Reference Beam Correction

For theoretical purposes we have made the simplification that light source drift between positive and negative encodings is negligible. This simplification is purely a theoretical device and in practice a reference measurement is acquired for and applied to each of the  $2N$  measurements. The purpose of light drift correction is to remove the bias due to  $R$  in the first term of equation 4.10. The simplest correction method for light source drift is to acquire reference spectra and divide the acquired spectra by the reference spectra before decoding. Assuming that the reference beam is acquired very quickly before or after the sample spectrum, light drift between sample and reference is negligible. The measured reference is represented by the diagonal matrix,  $R_r$ , and is the combination of the true light level values,  $R$ , and error terms, with entries

$$\{R_r\}_{i,i} = \alpha_r \{R + E_{P,r}\}_{i,i} + \{\mathbf{e}_r\}_i, \quad (4.21)$$

where  $E_{P,r}$  is the reference photon noise and  $\mathbf{e}_r$  is the instrument noise. The reference beam is designed to pass the maximum amount of light to the sensor, thus  $\{R\}_{i,i} = \text{var}(\{E_{P,r}\}_{i,i}) \gg \text{var}(\{\mathbf{e}_r\}_i)$ , moreover  $\{R\}_{i,i} \gg \text{var}(\{E_{P,r}\}_{i,i})^{1/2}$ . The total error variance in the reference beam is then well approximated by the photon noise,  $\sigma_{r,i}^2 = \alpha_r \langle r \rangle$ .

Application of the reference beam correction and then decoding gives

$$\begin{aligned} \hat{\mathbf{p}} &= \frac{1}{N} H \frac{1}{\alpha_r} R_r^{-1} [\alpha_s R H \mathbf{p}] + \mathbf{e}_t \\ &= \frac{\alpha_s}{\alpha_r} \mathbf{p} + \mathbf{e}_t, \end{aligned} \quad (4.22)$$

where  $\mathbf{e}_t$  is the total random error in the estimate, that is, the noise due to instrument and

photon effects combined with the reference beam noise. The total MSE after reference correction is

$$\begin{aligned}\epsilon_t &= \frac{1}{N} \text{var} \left( \frac{\mathbf{a}}{\mathbf{R}} \right) \\ &= \frac{1}{N \langle \mathbf{R}_r \rangle^2} \left( \sigma_a^2 + \frac{\langle \mathbf{a} \rangle^2}{\langle \mathbf{R}_r \rangle^2} \sigma_r^2 - 2 \frac{\langle \mathbf{a} \rangle}{\langle \mathbf{R}_r \rangle} \sigma_{a,r}^2 \right),\end{aligned}\quad (4.23)$$

where the second line in equation 4.23 uses the equation for the variance of the ratio of two variables with random error (Pearson, 1897),  $\langle \mathbf{a} \rangle$  is the expected value of the acquisition and  $\sigma_{a,r}^2$  is the covariance between the the sample spectra and the reference beam spectra. Substituting  $\sigma_a^2 = \alpha_s \langle r \rangle N \langle p \rangle + \sigma^2$ ,  $\sigma_r^2 = \langle \mathbf{R}_r \rangle = \alpha_r \langle r \rangle$  and  $\langle \mathbf{a} \rangle = \alpha_s \langle r \rangle N \langle p \rangle$  gives

$$\epsilon_t = \frac{1}{N (\alpha_r \langle r \rangle)^2} \left( \alpha_s \langle r \rangle N \langle p \rangle + \sigma^2 + \frac{\alpha_s^2 \langle r \rangle^2 N^2 \langle p \rangle^2}{\alpha_r^2 \langle r \rangle^2} \alpha_r \langle r \rangle - 2 \frac{\alpha_s \langle r \rangle N \langle p \rangle}{\alpha_r \langle r \rangle} \sigma_{a,r}^2 \right).\quad (4.24)$$

Only the reference beam is chopped (section 4.1.1), thus the raw measurements include light from both the sample and the reference beam. The reference beam measurements are easily obtained by subtracting the sample measurement from the raw measurement. This subtraction results in a covariance of  $\sigma_{a,r}^2 = -\sigma^2 - \alpha_s \langle r \rangle N \langle p \rangle$ . Substituting and rearranging, the MSE of the reference corrected estimates is

$$\epsilon_t = \frac{N \alpha_s \langle r \rangle \langle p \rangle + \sigma^2}{N \alpha_r^2 \langle r \rangle^2} \left( 1 + \frac{2N \alpha_s \langle p \rangle}{\alpha_r} \right) + \frac{N \langle p \rangle^2 \alpha_s^2}{\alpha_r \langle r \rangle \alpha_r^2}.\quad (4.25)$$

There is a factor of  $N$  on the top line in the brackets and in the second term. This factor of  $N$  corresponds to the same  $N$  that appeared in equation 3.53, which was considered disadvantageous. Note that the effect of the optical pathway is given by  $\alpha_s$  and  $\alpha_r$ , and that typically the light path of any reference beam system is designed to pass as much light as possible, thus  $\alpha_r \gg \alpha_s$ . To a close approximation

$$\epsilon_t \approx \frac{N \alpha_s \langle r \rangle \langle p \rangle + \sigma^2}{N \alpha_r^2 \langle r \rangle^2}.\quad (4.26)$$

The decoding performs an averaging operation of the noise, so the average MSE is  $\epsilon = \epsilon_t$ . Comparing equation 4.26 to equation 4.15 we see that the error due to the bias is removed and the error is reduced by a factor of  $\alpha_r^2 \langle r \rangle^2$ . The mean squared signal is also reduced by the factor of  $\alpha_r^2 \langle r \rangle^2$  and so the overall signal to noise ratio is not reduced by this factor. The removal of the bias term is significant if the drift is significant. If the additive noise variance,  $\sigma^2$ , is larger than the multiplicative noise, then the noise is reduced by the Hadamard multiplexing. However if  $N \alpha_s \langle r \rangle \langle p \rangle \gg \sigma^2$  then the noise is not reduced, so in such cases multiplexing is not necessary. In situations where both may occur then

equation 4.26 may be considered a preferential noise reduction, which is demonstrably useful when the main noise source is indeterminate or varying.

For G-matrix multiplexing with reference correction it follows that the total noise variance, hence the average MSE is twice that of the H-matrix

$$\epsilon_t \approx \frac{2N\alpha_s \langle r \rangle \langle p \rangle + 2\sigma^2}{N\alpha_r^2 \langle r \rangle^2}. \quad (4.27)$$

For pointwise encoding the total MSE per pixel is

$$\epsilon_{t,j} \approx \frac{\alpha_s \langle r \rangle_j p_j + \sigma^2}{\alpha_r^2 \langle r \rangle_j^2}. \quad (4.28)$$

The average MSE is then the average of the  $\epsilon_{t,j}$ , thus

$$\epsilon = \frac{1}{\alpha_r^2 N} \sum_j \left( \frac{\alpha_s p_j}{\langle r \rangle_j} + \frac{\sigma^2}{\langle r \rangle_j^2} \right). \quad (4.29)$$

In equation 3.43 we made an approximation that the mean of a product is the product of the means. This approximation is not valid for ratios so equation 4.29 is the simplest form of the average MSE for reference corrected pointwise measurement.

### 4.1.5 Reference Object Correction: the White Tile Effect

The optics used in the multiplexing produce spatial and spectral inhomogeneities that cause systematic error. Furthermore the error was assumed constant over the spatial and spectral dimensions. In practice the error due to the optics is not constant but rather varies in space and wavelength, however the optics effect is multiplexed with the pixels and subsequently recovered intact at decoding thus the above theoretical results hold. In precision measurement we want to obtain the variation due to the sample only and any systematic error must be corrected for. The Standard practice in reflectance spectroscopy is to measure a white reference tile with very high diffuse reflectance. We use a Spectralon reference white tile (Labsphere, North Sutton, NH, USA) which the manufacturers claim to have the highest known diffuse reflectance in the visible and near infrared domain. Furthermore the Spectralon white tile has a very flat reflectance profile both spatially and spectrally in the spectral domain of interest (approximately 400–1700 nm).

The purpose of white tile reference correction is to correct for optical effects not directly due to the sample. A spectroscopic measurement obtains the mathematical product of the attenuation due to each optical substance and boundary encountered by the light, including the optical glass, mirrors, the light source, the spectrometer diffuser, the sensors

and even air as well as the sample. Replacing the sample with a reference tile and acquiring spectra we essentially assume that the attenuation due to the reference tile is known, to which sample reflectances are compared. Taking the ratio of the sample spectra to the white tile spectra removes the effect due to the optics.

Recall that the reconstructed image pixels of an object with reference beam correction,  $\hat{\mathbf{p}}$ , are (equation 4.22)

$$\hat{p}_i = \frac{\alpha_{s,i}}{\alpha_r} p_i + e_{t,i}.$$

The reference white image,  $\hat{\mathbf{p}}_W$ , is accordingly

$$\hat{p}_{W,i} = \frac{\alpha_{s,i}}{\alpha_r} \rho_W + e_{t,W,i}. \quad (4.30)$$

where  $\rho_W$  is a real number less than but close to 1 that represents the reflectance of the tile. In practical situations it is nearly impossible to accurately assess  $\rho_W$  with high precision, so we use the pragmatic approximation that the reference tile is a perfect reflector, i.e.  $\rho_W = 1$ . Taking the pixelwise ratio of the pixel to reference tile measurements gives

$$\frac{\hat{p}_i}{\hat{p}_{W,i}} = p_i + e_{C,i} \quad (4.31)$$

where the  $e_{C,i}$  are the noise errors after white tile correction.

White tile correction, like the reference beam correction, removes a bias to make the estimate of  $\mathbf{p}$  more accurate. However such corrections also increase the noise variance. Returning to the assumption that the  $\alpha_{s,i}$  have the same value, the noise MSE for the reference corrected pixel values is (equation 4.26)

$$\epsilon_t \approx \frac{N \alpha_s \langle r \rangle \langle p \rangle + \sigma^2}{N \alpha_r^2 \langle r \rangle^2}.$$

Immediately we can see that the MSE for the reference corrected white tile measurements is

$$\epsilon_{t,W} \approx \frac{N \alpha_s \langle r \rangle + \sigma^2}{N \alpha_r^2 \langle r \rangle^2}. \quad (4.32)$$

We observe that  $\langle p \rangle < 1$ , thus  $\epsilon_{t,W} > \epsilon_t$  in the presence of Poisson noise. This is a sensible result as the white tile diffusely reflects more light to the sensor than any other known sample.

The average MSE due to the white tile correction, using Pearson's formula and using

the fact that white tile measurements are independent of the image measurement, is

$$\begin{aligned}\epsilon_C &= \text{Var} \left( \frac{\hat{p}}{\hat{p}_W} \right), \\ &= \frac{1}{\langle \hat{p}_W \rangle^2} \left( \epsilon_t + \frac{\langle \hat{p} \rangle}{\langle \hat{p}_W \rangle} \epsilon_{t,W} \right).\end{aligned}\tag{4.33}$$

Using the above expressions for the MSEs,  $\langle \hat{p} \rangle = \alpha_s \langle p \rangle / \alpha_r$  and  $\langle \hat{p}_W \rangle = \alpha_s / \alpha_r$ ,

$$\begin{aligned}\epsilon_C &\approx \frac{\alpha_r^2}{\alpha_s^2} \left( \frac{N \alpha_s \langle r \rangle + \sigma^2}{N \alpha_r^2 \langle r \rangle^2} + \langle p \rangle^2 \frac{N \alpha_s \langle r \rangle + \sigma^2}{N \alpha_r^2 \langle r \rangle^2} \right), \\ &= \frac{N \alpha_s \langle r \rangle \langle p \rangle + \sigma^2}{N \langle r \rangle^2 \alpha_s^2} + \langle p \rangle^2 \frac{N \alpha_s \langle r \rangle + \sigma^2}{N \langle r \rangle^2 \alpha_s^2}.\end{aligned}\tag{4.34}$$

In equation 4.34 we see the multiplexing mitigation of the effect of additive noise. However there is a sharp increase in noise variance over the reference corrected measurements due to the optics variable,  $\alpha_s^2$ , appearing on the bottom line of both terms. The light intensity term,  $\langle r \rangle$ , is typically large, so the increase in variance is counteracted somewhat by the multiplication of  $\alpha_s^2$  by  $\langle r \rangle^2$ . Further compounding the noise increase is a near doubling of noise variance seen by the presence of two terms of similar magnitude. The theoretical analysis leads to the common sense conclusions that, apart from the multiplex advantage, more light from the light source and less attenuation due to the optics ultimately leads to better signal quality. Equation 4.34 takes the interpretation further in showing how poor light levels and high light attenuation not only causes a low signal at capture, but causes high noise levels when all the necessary corrections are applied.

## 4.2 Testing Methodology

To test the hardware we take images and examine:

1. the accuracy and repeatability of measurement and
2. the usability of the spectra produced.

Hadamard multiplexing is assumed accurate thus for Hadamard multiplexing we only assess the SNR per wavelength. For compressed sensing the accuracy and precision varies with the number of measurements so we need to test for the similarity to the true values as well as the SNR. In practical cases it is typically not possible to know the true values so we must use the best estimate of the true values possible. Since the Hadamard multiplexing is assumed accurate we use the average of a number of Hadamard acquired images as the best estimate of the true values.

SNR is simply the square root of the mean squared signal over the variance. To characterise the SNR for hyperspectral imaging we:

- capture a set of images;
- compute the mean and variance for each pixel at each wavelength and
- compute the SNR hyperspectral image.

Clearly for a hyperspectral system there is a SNR value associated with each pixel at each wavelength and we must reduce the rich information down into an easily interpretable form. The imager performs the multiplexing operation simultaneously to all the wavelengths measured, thus spectral variation in the SNR and SNR boost are informative of the nature of the noise per wavelength. We take the mean SNR over the pixels to obtain a single SNR spectrum. Computing an average SNR spectrum is consistent with the concept of A-optimality (section 3.2.1). The multiplexing operates over the pixels, and we want to quantify on average the power of multiplexing to improve signal quality.

When a spectral image is obtained that is sufficiently accurate and precise then it is usable. Spectral images that are usable can be processed to yield a discrete segmentation or continuous range of values that represent some characteristic of the object. Spatial representations, such as the mean image, principal components over the wavelengths or multivariate calibrations, are interesting when there is spatial variation in the object. The processed versions of the hyperspectral image typically have much lower dimensionality so it is a simple matter to view and qualitatively assess gross features. Spatial representations are further transformable. Simple thresholding or grey scale slicing, or more complex processing techniques such as clustering, can be employed to return a classification of pixels. Calibration techniques such as regression on the principal components against some reference value return a single continuous prediction image. In this chapter we are interested in simple examples which demonstrate that the hyperspectral imager provides useful data.

### 4.2.1 G-matrix (Complement S-matrix) Multiplexing

In addition to H-matrix complement encoding we test G-matrix encoding via S-matrix complement encoding (see Section 3.1.1 for an overview of the matrices). Recall from Section 4.1.2 that complement coding, in conjunction with the source modulated multiplexing optics, facilitates the removal of background illumination effects from the measurements. G-matrix complement encoding is implemented by acquiring data according to  $S$  and by  $J - S$ , where  $G = S - (J - S) = 2S - J$  and the two encodings acquire complement information. Each row of  $S$  is wrapped and projected onto the sample, following

immediately by the corresponding row of  $J - S$  so that drift effects between acquisitions are minimised. The acquired data are reference beam corrected and subtracted to give

$$\begin{aligned}\mathbf{a} &= (J - S)\mathbf{p} - S\mathbf{p} + \mathbf{e}_t, \\ &= (J - 2S)\mathbf{p} + \mathbf{e}_t, \\ &= G\mathbf{p} + \mathbf{e}_t,\end{aligned}\tag{4.35}$$

where  $\mathbf{e}_t$  is the total noise. Previously  $2S - J$  was referred to as the total encoding (Streeter *et al.*, 2007, 2008b); here we correctly identify the total encoding with  $G$ . If  $G$  is order  $N - 1$ , then the estimate from decoding is

$$\begin{aligned}\hat{\mathbf{p}} &= \frac{2}{N}S\mathbf{a} \\ &= \mathbf{p} + \frac{2}{N}S\mathbf{e}_t.\end{aligned}\tag{4.36}$$

From Sections 3.2.1 and 3.2.3 we expect a reduction in additive random noise variance of  $2/N$  and a doubling of the Poisson noise variance. The overall effect on the SNR depends on the relative contributions of the two noise sources.

### 4.2.2 Compressed Sensing

For compressed sensing (CS) we used the same encoding pattern  $\Phi$  of randomly entered ones and negative ones in each image acquisition (Streeter *et al.*, 2008a). Complement encoding is used to acquire the data:  $\Phi$  is split into complement positive and negative parts  $\Phi^+$  and  $\Phi^-$  with entries that consist of ones and zeros. We acquire

$$\begin{aligned}\mathbf{a}^+ &= \Phi^+\mathbf{p}, \\ \mathbf{a}^- &= \Phi^-\mathbf{p},\end{aligned}\tag{4.37}$$

and take the difference which gives

$$\mathbf{a} = \mathbf{a}^+ - \mathbf{a}^- = \Phi\mathbf{p}.\tag{4.38}$$

To compare the Hadamard and CS encoding on a fair basis we use  $\Psi = H$ , the Hadamard matrix, as the compression basis. This choice is further motivated by the fact that if one concatenates the rows of an image, as we have, then strong almost square shaped periodicity occurs. Periodic information is exactly the type of information that the Hadamard transform can represent in a sparse manner. For the CS decoding the following was carried out:

1. for  $l$  from 50 to 256,

2. estimate  $\hat{\mathbf{p}}$  from  $l$  measurements.

We ran the CS decoding as per equation 2.24, i.e. non-noise aware, and with the noise aware option (Tsaig and Donoho, 2006). The Sparselab package (see Section 2.4) has a basis pursuit denoising option which is invoked by setting  $\delta = 1$  in equation 2.25. The reference beam correction reduces the magnitude of the data passed into the basis pursuit reconstruction, so noise aware CS used a noise parameter of  $\delta = 0.001$ . This noise parameter term appears somewhat small, but is further motivated by the expected noise level with respect to the reference corrected signal. After reference beam correction and subtraction of the negative from the positive encoding this noise parameter is of the same order of magnitude as the measurements.

The SNR was used to examine precision and the RMS difference from the best estimate of the true signal used to examine accuracy. For precision the mean and standard deviation for each type of imaging were taken and the SNR hyperspectral images computed. The ratio of SNRs between the Hadamard to pointwise and the CS to pointwise were computed. The mean SNR over the pixels were taken to give a single average SNR boost spectrum for each case. For accuracy the mean of the Hadamard images was used as the best estimate of the true pixel values. The difference between the CS and mean Hadamard reconstructions were computed to obtain a hyperspectral difference image. For each number of measurements used in reconstruction the RMS error was computed over the acquisitions and the mean RMS error was computed and plotted.

### 4.2.3 Objects Imaged

A set of objects were imaged to provide interesting hyperspectral images. In some cases the objects are intended to simply show upon visual inspection that the imager produces a sensible spatial pattern, that is, the image looks correct. Other objects were fabricated to provide contrasting signals so that we can observe the spectral shape. Segmentation of the hyperspectral images by simple chemometric techniques then demonstrates the usability of the spectra produced by the imaging system.

A spatial grid pattern was generated and is shown in Figure 4.5. The image shown is slightly larger than the view area used in the hyperspectral imaging. However in the hyperspectral image a white square in the top lefthand region is approximately  $6 \text{ mm}^2$  where in the bottom righthand corner a square is less than  $3 \text{ mm}^2$ . The grid pattern is for performing imaging at different spatial resolutions. The pattern is comprised of a square wave sequence that varies with increasing rate from left to right and from top to bottom. Where two peaks in the sequences intersect then a white region results, when two troughs intersect then a black region results and where a trough meets a peak then a 50% grey region results. A given imaging resolution that is able to resolve the top left may produce

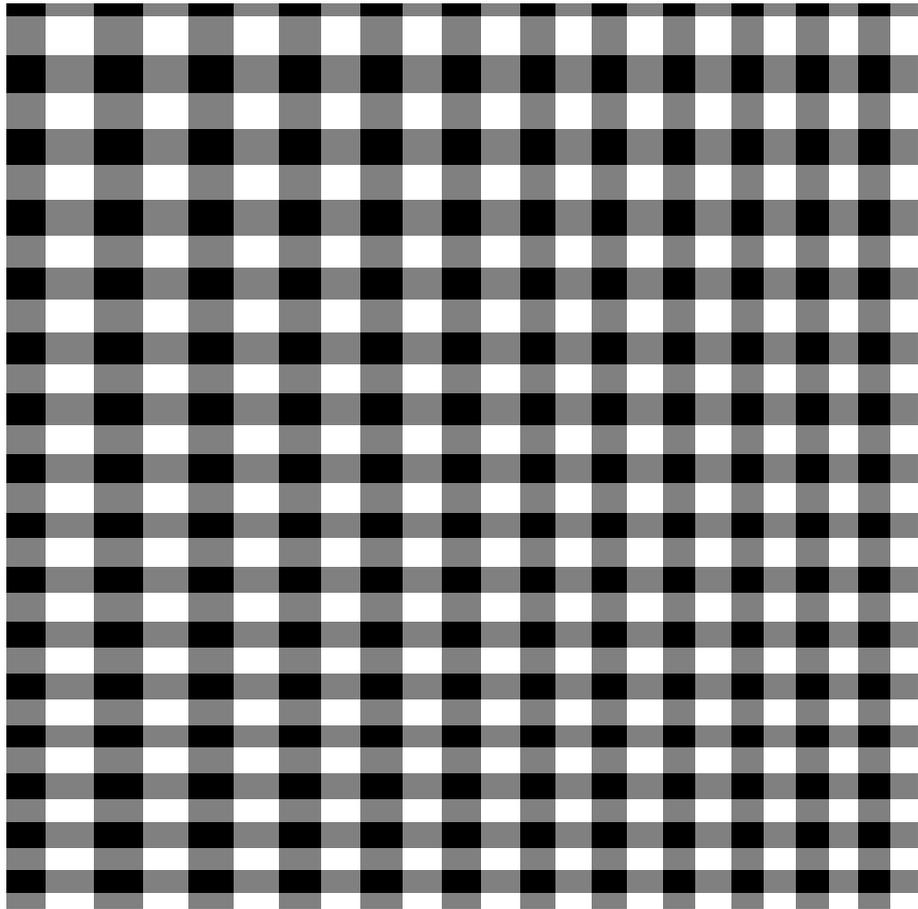


Figure 4.5: The testpattern for examining spatial variation in the reconstruction. The rate of change of the bright-dark patterns increases from the top left to the bottom right.

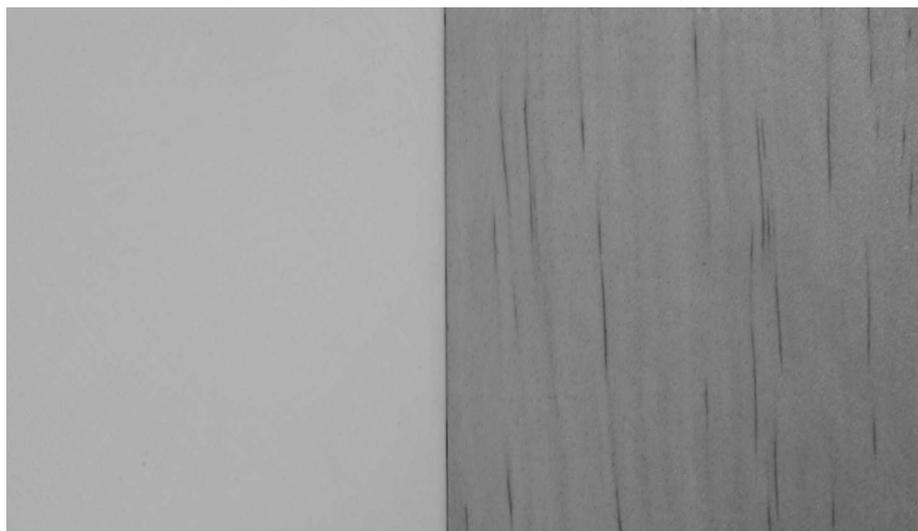


Figure 4.6: The wood and acrylic object. The acrylic is on the left and the wood on the right.

aliasing at the bottom right. Furthermore if the imaging method is valid then the pattern, or some aliased corruption of the pattern, should be clearly apparent in reconstructed image data. The spatial grid pattern was printed using a typical laser jet printer on printer paper and glued onto a cardboard back using PVA. Also a fully black image was printed from the same printer to test the ink. The ink exhibited reasonably homogeneous absorption over the entire spectral region of the spectrometer (400–1700 nm wavelength) with no obvious peaks.

Figure 4.6 shows the wood and acrylic object used by Streeter *et al.* (2007) and imaged using G-matrix encoding. Streeter *et al.* (2007) used photographic white paper on an aluminium backing as the reference white. While not ideal the white paper provided the means to correct for spatial inhomogeneities, albeit without the optimised reflectance of a Spectralon tile. The wood is pine (*pinus radiata*) and the acrylic is the material commonly used in engineering workshops. The wood is approximately 20 mm thick and the acrylic approximately 3 mm thick. The two materials are held together by PVA glue, with another piece of pine glued to the wood behind the acrylic, onto which the acrylic is glued. Acrylic was chosen for this object because of its distinct spectral absorption regions in the NIR. In contrast wood was chosen because it is a highly inhomogeneous material that can be expected to have a distinct general shape but with considerable variability. The wood used in this sample is quite dry, so distinct water absorption bands should not be obvious.

Figure 4.7 shows the polystyrene object with diagonal inlays of Rimu wood (*dacrydium cupressinum*) strips. This object was used in Streeter *et al.* (2007, 2009). In the latter publication imaging was performed with H-matrix encoding. Both the wood and the polystyrene are approximately 15 mm thick and glued directly onto a cardboard back. The shape of this object is to show spectral imaging of an object with a more interesting spatial pattern than simply two halves. Polystyrene is highly particulate so spectra produced are expected to have a high degree of scattering effect. Also the material was deliberately obtained from an engineering workshop where a reasonable chance of contamination is possible by dust or other substances, so ‘interesting’ spectra are expected. Similarly the Rimu wood was from a discarded piece of furniture, although surfaces without paint or polish were used.

In Streeter *et al.* (2008b) a sample with acrylic on the left hand side and white cardboard on the right hand side was used to demonstrate the usability of the spectra and examine the SNR boost of the multiplexing with reference beam correction. Forty images were taken of the object over five days to allow for the possibility of changes in the imager to occur and to examine if the reference correction reduced these changes. The Spectralon tile was used as the white reference. G-matrix encoding was employed as the imaging method. A Figure of this sample is not included because in the visible domain it appears as a plain white object with a crack down the middle. To the eye the cardboard

and the acrylic are distinguishable, however a colour machine vision system would have considerable difficulty separating the two substances. Cardboard, like wood, is likely to have considerable chemical variation. While the white cardboard board and acrylic look similar in the visible domain, in the NIR there are strong differences and discrimination is a simple matter.

The acrylic and printed circuit board (PCB) object in Figure 4.8 was fabricated as an alternative to the acrylic and wood object and was used in Streeter *et al.* (2009). The sample consists of the same type of acrylic used in the acrylic and wood sample with PCB fibreboard fastened with electrical tape. When imaging the object the tape was kept out of the optical pathway. The PCB board is semi-transparent and yellowish to the eye. The spectral effect of the PCB was largely unknown at the time of fabrication. However general attenuation of the light was expected.

The Spectralon white tile was imaged to act as a reference white object. Pointwise division of sample images by the Spectralon images corrects for unwanted variability due to the imager. Also the images of the Spectralon tile indicate the spatial variation due to the imager. We use the Spectralon tile for performing SNR measurements on the system. For the G-matrix imaging we use a 255 order matrix derived from a cyclic SRS S-matrix wrapped to acquire  $15 \times 17$  pixel images. For H-matrix we use a 256 order Sylvester construction matrix wrapped to acquire  $16 \times 16$  pixel images. For each SNR boost experiment pointwise images were acquired concurrently with the multiplexed images of the same dimension and resolution. The white tile has the highest diffuse reflection of any object available to us. Thus the light level from the tile, hence the photon noise, is greatest when using the Spectralon tile. The SNR of imaging with the white tile will be greatest with respect to additive signal independent noise, but the SNR boost reduction due to multiplicative signal dependent noise is maximised. Thus the SNR of the white tile images provide the most interesting and informative SNR data.

## 4.3 Results and Discussion

### 4.3.1 Images and Spectra

Images of the spatially varying pattern in figure 4.5 are shown in Figure 4.9. Both images were taken with Hadamard H-matrices and are displayed at 1178 nm. The pattern is homogeneous over the wavelengths where there is sufficient SNR. Both images were filtered in the spectral dimension to reduce noise. There is an edge effect on the top and bottom right due to the spectrometer not quite capturing light from this region. The images have been white corrected which exacerbates the error in the edge effects considerably where the signal, hence SNR, is small in magnitude. The imaging system was designed for

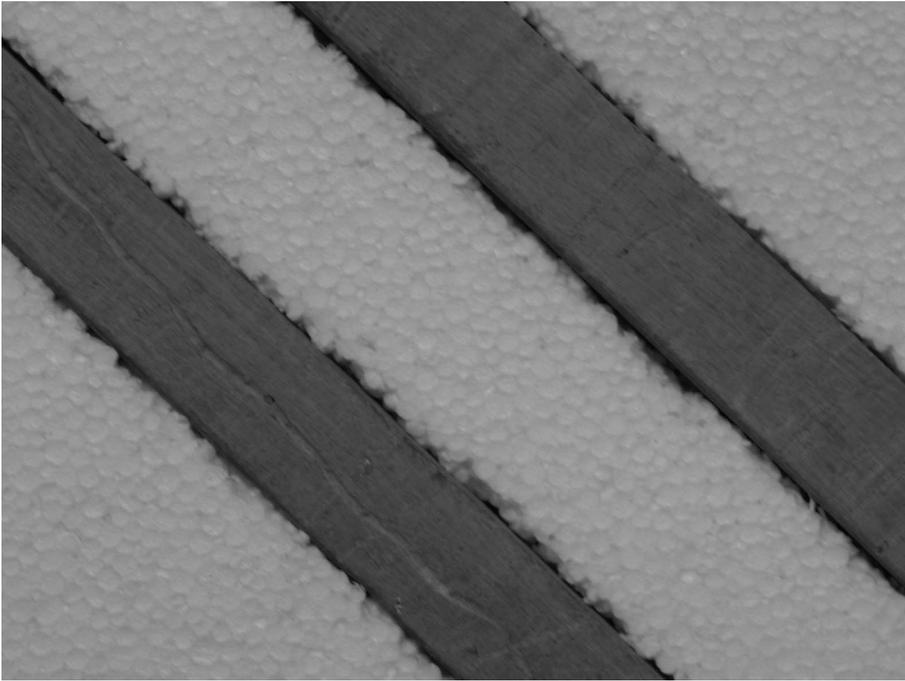


Figure 4.7: The wood and polystyrene object.

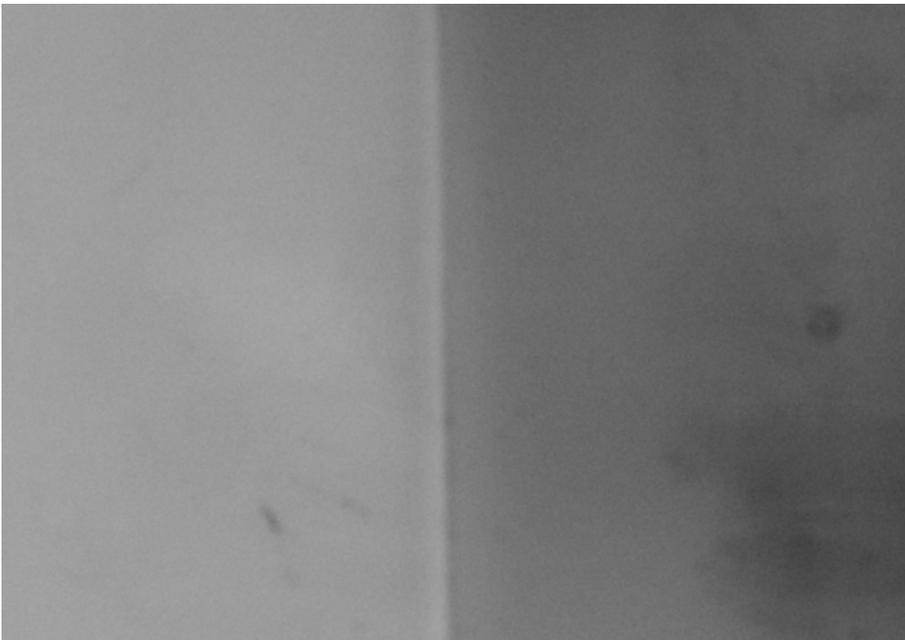


Figure 4.8: The acrylic and PCB object. The acrylic is on the left and the PCB covering the acrylic on the right. Marks are imperfections such as scuffs on the material.

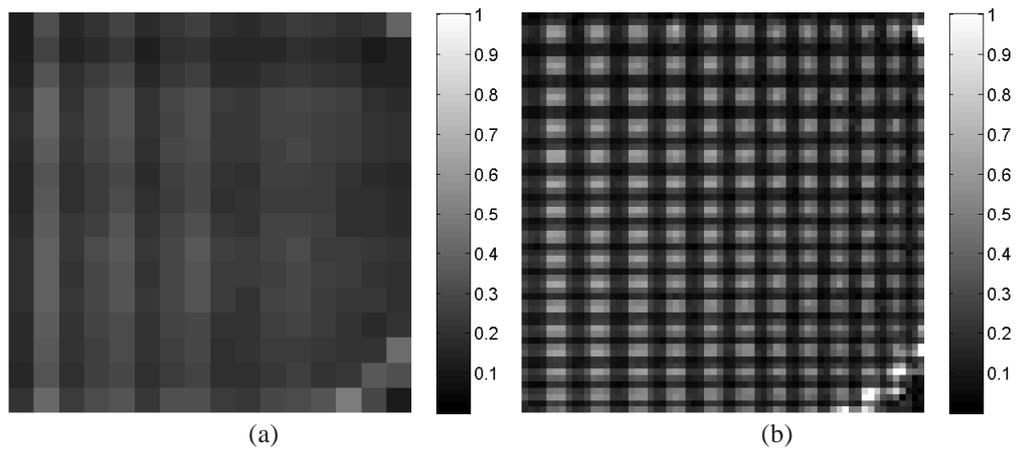


Figure 4.9: Images of the test pattern. a) A  $16 \times 16$  pixel image and b) a  $64 \times 64$  pixel image. Both images were of an approximately  $100 \times 100$  mm square region on the object.

imaging a round object centred in the middle of the field of view (see the next chapter), so the corner edge effects are not significantly detrimental for practical applications. The left hand image has resolution of  $16 \times 16$  pixels over approximately  $100 \times 100$  mm. This corresponds to a pixel size of about  $6 \times 6$  mm, far too low resolution to resolve the pattern. Indeed visual inspection immediately confirms that the pattern is severely aliased. On the left and to a lesser degree at the top a square pattern is seen, but is completely lost at the right hand bottom. The right hand image has a resolution of  $64 \times 64$  pixels over approximately  $100 \times 100$  mm, resulting in pixel size of about  $1.6 \times 1.6$  mm. The square pattern of varying rate of change is clearly visible. Some aliasing is present, especially towards the right and bottom of the image.

With the available equipment the lower resolution image took 2.5-3 minutes to acquire and another 2.5-3 minutes for the accompanying white reference tile image, for a total of 5 minutes. The higher resolution image took over 40 minutes, 80 minutes including the white tile image. The acquisition time for one spectrum is the length of time from the computer command to change the pattern until the image completely forms on the DMA plus the acquisition time of the spectrum for that pattern, in total about 8/25s. During the spectral acquisition four spectra are taken and averaged to reduce noise which represents about half the acquisition time. Reducing the number of spectra would reduce the acquisition time but will also increase the noise. DMA driver hardware is available off the shelf (albeit expensive) with a frame rate of 100 fps, that potentially can reduce the acquisition times drastically. With DMA frame rate of 100 fps data acquisition is limited by the capture rate of the spectrometer.

Figure 4.10 shows spectra from the tenth row from the top of the image of the wood and acrylic object. The object was imaged with G-matrix encoding wrapped into a  $15 \times 17$  pixel pattern. Some of the spectra have reflectance substantially greater than

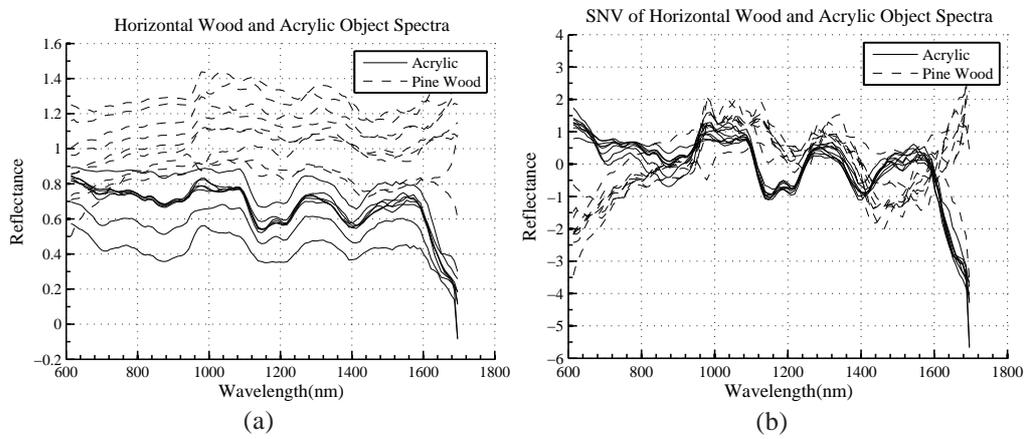


Figure 4.10: Acrylic and wood object imaged using G-matrix encoding. a) Reference corrected spectra and b) the standard normal variate transform of the same spectra.

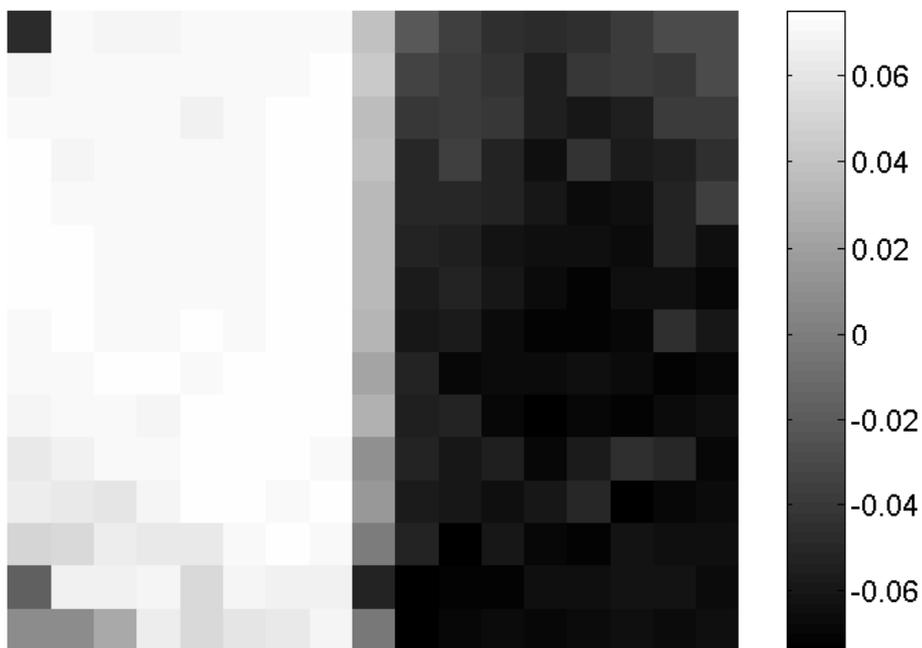


Figure 4.11: The first principal component of the image of the wood and acrylic object.

unity which is due to the use of photographic paper in this particular instance. Regardless there are obvious differences in the spectral shapes of the two substances. The differences are emphasised in the SNV transformed spectra. The acrylic spectra are quite uniform in shape whereas the wood spectra show considerable variation due to chemical variation in the wood. The familiar water absorption bands are not present in the wood due to the wood being sufficiently dry. The acrylic has the usual absorption bands around 1150 nm, 1200 nm, 1400 nm and towards 1700 nm. A mixed boundary spectrum is apparent, particularly in the regions around 1150 nm and 1200 nm.

Figure 4.11 shows the first principal component of the SNV spectra from the wood and acrylic hyperspectral image. The difference between the two substances is clearly visible. There are edge effects in the top left and bottom left corners due to the positioning of the spectrometer. Mixed spectra down the centre of the image are apparent and correspond to where the pixels straddle both the wood and the acrylic.

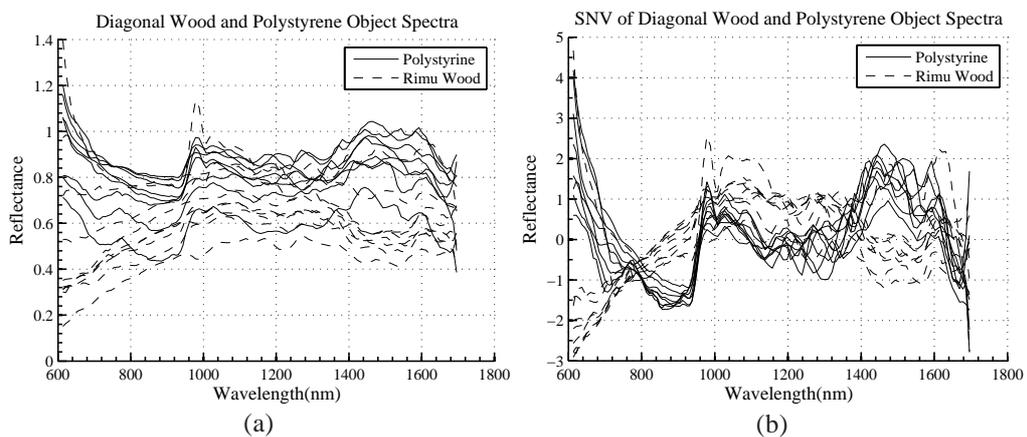


Figure 4.12: Wood and polystyrene object imaged using G-matrix encoding. a) Reference corrected spectra and b) the standard normal variate transform of the same spectra.

Figure 4.12 shows the spectra from the diagonal wood and polystyrene object (Figure 4.7) acquired using G-matrix encoding and reference corrected with the photographic white paper. The encoding was wrapped into a  $15 \times 17$  pixel pattern for imaging. There is some reflectance greater than unity due to the white paper not having optimal reflectance. The spectra are messy due to contamination in the sample and possibly due to scattering in the polystyrene. The SNV of the spectra clarifies the grouping between the two groups of spectra somewhat, but the noise is still a dominant feature. Figure 4.13 shows the second principal component of the wood and polystyrene image. The diagonal wood inlays are clearly visible, appearing as dark regions. The fact that the second principal component and not the first describes the between group variation is most likely due to both random and non-random error in the spectra.

Figure 4.14 shows spectra from the eighth row of image thirty of the white cardboard

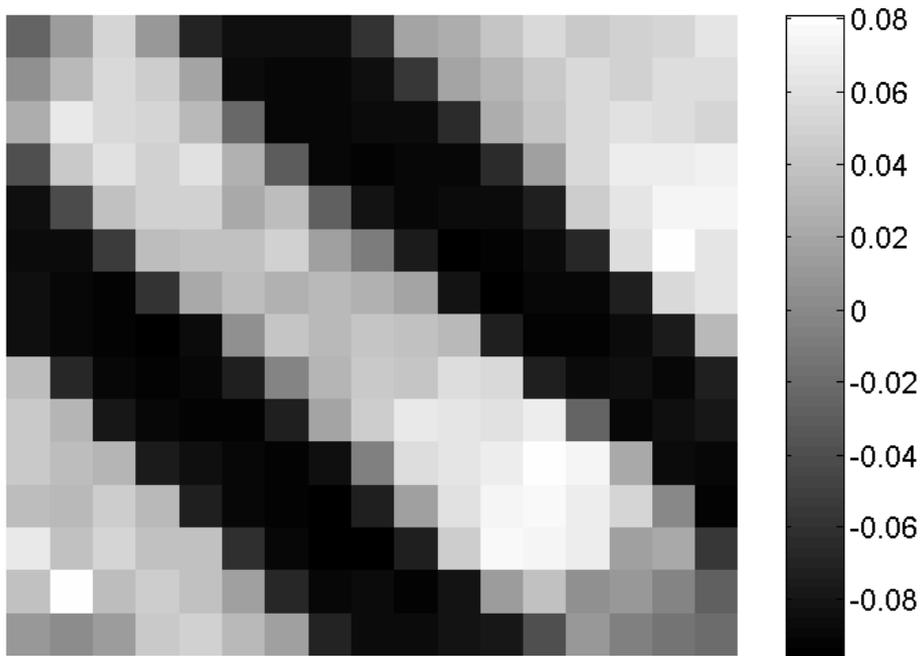


Figure 4.13: The second principal component of the image of the wood and polystyrene object. The imaging was performed with a 255 order G-matrix.

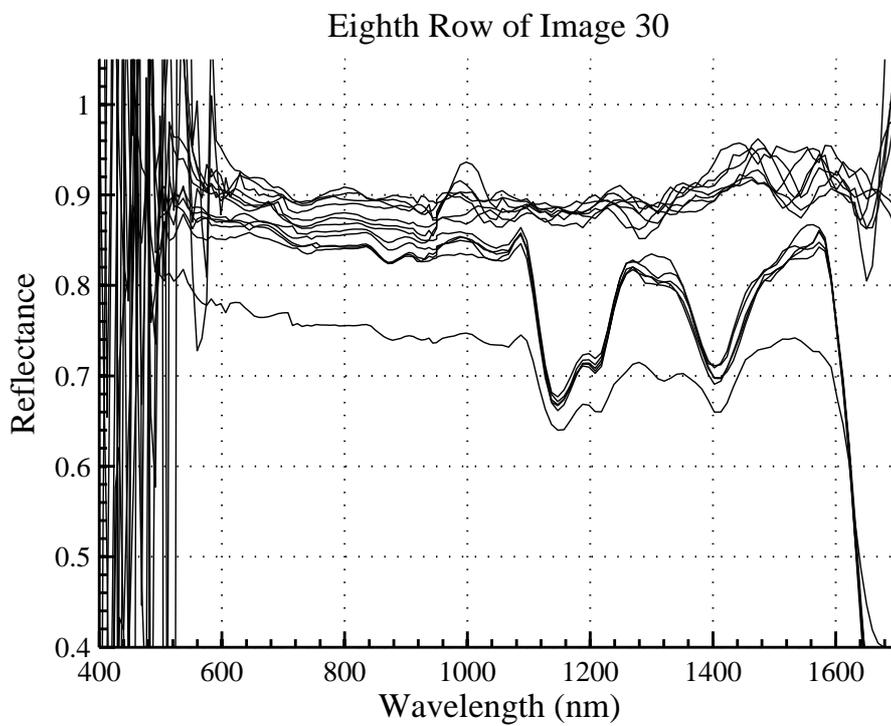


Figure 4.14: Spectra from the eighth row of an acrylic and white cardboard object.

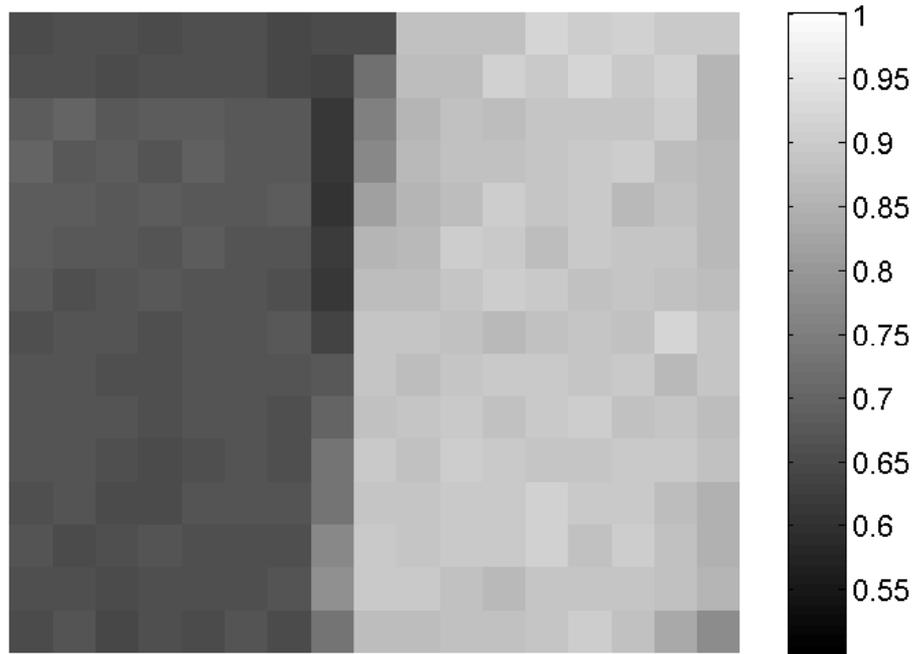


Figure 4.15: The thirtieth spectral image of the acrylic and cardboard object at 1150 nm.

and acrylic object. In this Figure we show the full spectral range of the spectrometer to display the severity of the noise in the extreme regions, particularly in the visible domain. Both substances have high reflectance so the reflectance axis starts at 0.4 to show clearly the detail of the spectral shape. The spectra of the acrylic are easily identifiable from the absorption bands around 1150 nm, 1200 nm, 1400 nm and towards 1700 nm. The cardboard spectra are predominately flat but have some spectral inhomogeneity. The spectrum with gross low reflectance is from an edge pixel. All reflectances greater than unity are due to noise. Figure 4.15 shows the spatial variation of the same spectral image at 1150 nm. The two substances are clearly distinguishable with acrylic on the left and cardboard on the right. The object was not perfectly aligned vertically which is seen as the slight diagonal lean of the boundary.

Figure 4.16 shows spectra from images taken using H-matrix encoding. The image resolution is  $16 \times 16$  pixels over approximately  $100 \text{ mm}^2$ . The spectra are from the eighth row of each image. These spectra were white corrected using the Spectralon tile. The wood spectra are easily identified by the lower overall reflectance and the water absorption bands about 1200 nm and 1450 nm. Comparing the wood and polystyrene object spectra with that of Figure 4.12 the increase in water features in the wood spectra is apparent. A period of some months transpired between the acquisition of the G-matrix image and the H-matrix image, during which the object was stored near a window in an office where it was undeliberately exposed to moisture. Regardless the difference in the spectral shapes are interesting. The acrylic and PCB object spectra are distinguishable by

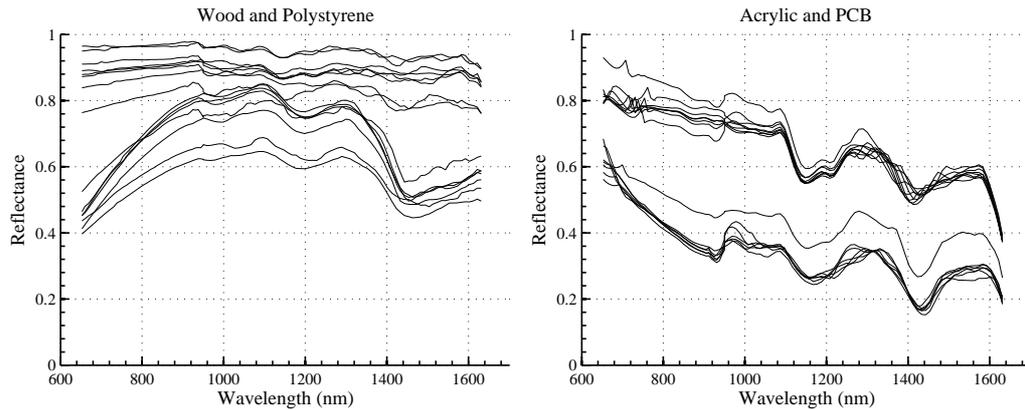


Figure 4.16: Spectra acquired by H-matrix encoding. The wood and polystyrene object is on the left and the acrylic and PCB fibreboard object is on the right.

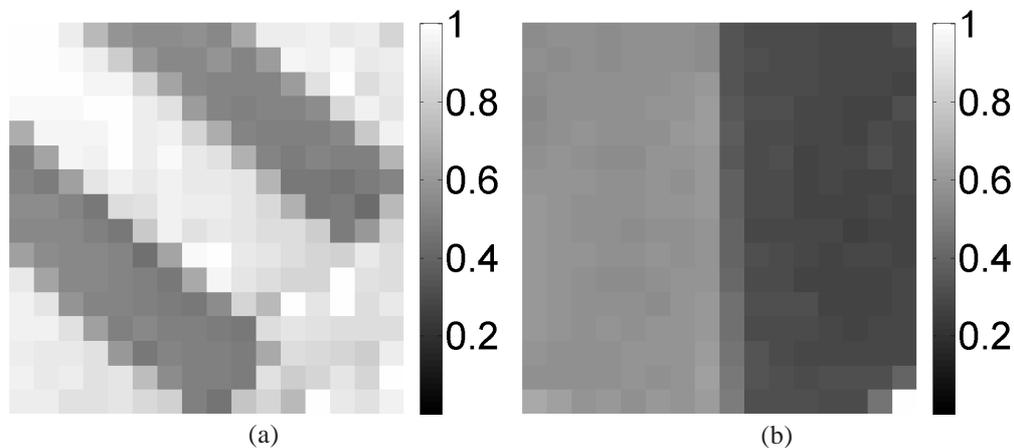


Figure 4.17: a) The wood and polystyrene object at  $1533\text{ nm}$  imaged using H-matrix encoding. b) The PCB and fibreglass object.

gross absorption, where the PCB attenuates the light travelling too and from the acrylic. Some alteration of the spectra shape due to the PCB fibreboard is seen.

Images acquired using the H-matrix encoding are in Figure 4.17. The wood and polystyrene object is on the left with the diagonal dark regions due to the wood. Here the wood and polystyrene object is positioned so that the end of the wood strips are in the field of view. The acrylic and PCB object is on the right. The PCB fibreboard region is apparent as the darker region on the right side of the image. In both images there is an edge effect in the bottom right hand corner.

### 4.3.2 Signal to Noise Ratios and Noise Features

Figure 4.18 shows the SNR for G-matrix encoding, G-matrix with reference correction per acquisition and G-matrix encoding with reference correction by the average reference

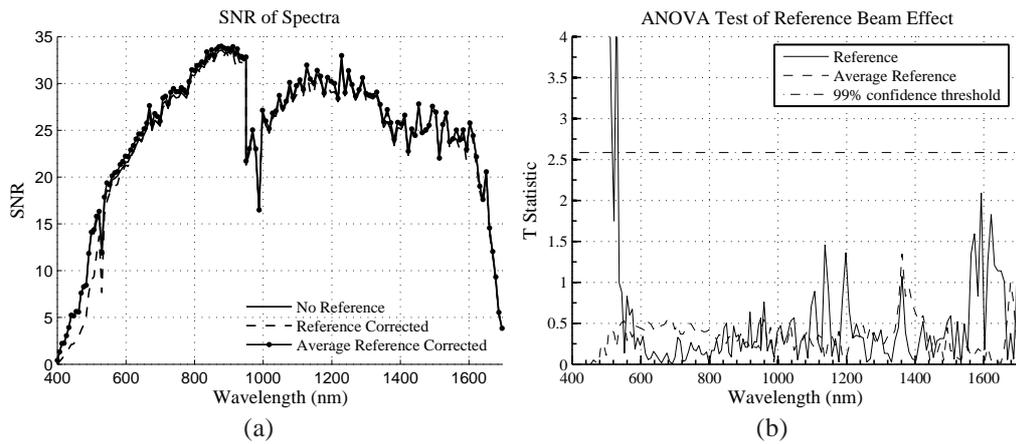


Figure 4.18: a) SNRs for the G-matrix encoding with no reference correction, correction per acquired spectrum and the average reference spectrum over an acquisition period for the entire encoding matrix. b) Significance test of reference correction using the of the reference measurements and the per sample reference measurements.

spectrum for each encoding period. An ANOVA test (Ott and Mendenhall, 1985) for the difference between the treatments with a 99% confidence threshold shows no significance above 550 nm. The T-statistic (the square root of the F-statistic used in ANOVA) is shown as we compared the SNR of each reference correction method to the non-corrected SNR separately in a two treatment manner. Below 550 nm the reference correction causes a significant degradation in the SNR. The precise reason why the SNR decreases is unknown. Regardless the low wavelength region corresponds to where the SNR is too poor for any use of the data. The rest of the spectral region shows no significance in the change of SNR due to reference correction. The lack of change indicates two things, 1) that the light source is stable during the data acquisition, so that the bias due to light source drift is small, and 2) that the reference beam is sufficiently intense, thus the noise level is not significantly increased by the correction.

Figure 4.19 shows the SNR boost for G-matrix encoding and for G-matrix encoding with reference correction. The smallest SNR boost around 750 – 800 nm, where the light from the source, hence the photon noise, is greatest. The boost reaches the theoretical maximum of  $\sqrt{255/2} = 11.29$  at the lowest and highest wavelengths where the light from the source is least and the boost is most needed. SNR boost values exceeding 11.29 are ascribed to random effects. The sharp discontinuity near 950 nm occurs at the boundary between the two types of sensor in the spectrometer.

Figure 4.20 shows the average SNR for H-matrix multiplexed imaging of a Spectralon white tile. The SNR is much greater than that seen in Figure 4.18 for G-matrix imaging, but the G-matrix images were white reference corrected where these are not, so a direct comparison is unfair. Below we compare SNR boosts which is a fair test. The H-matrix

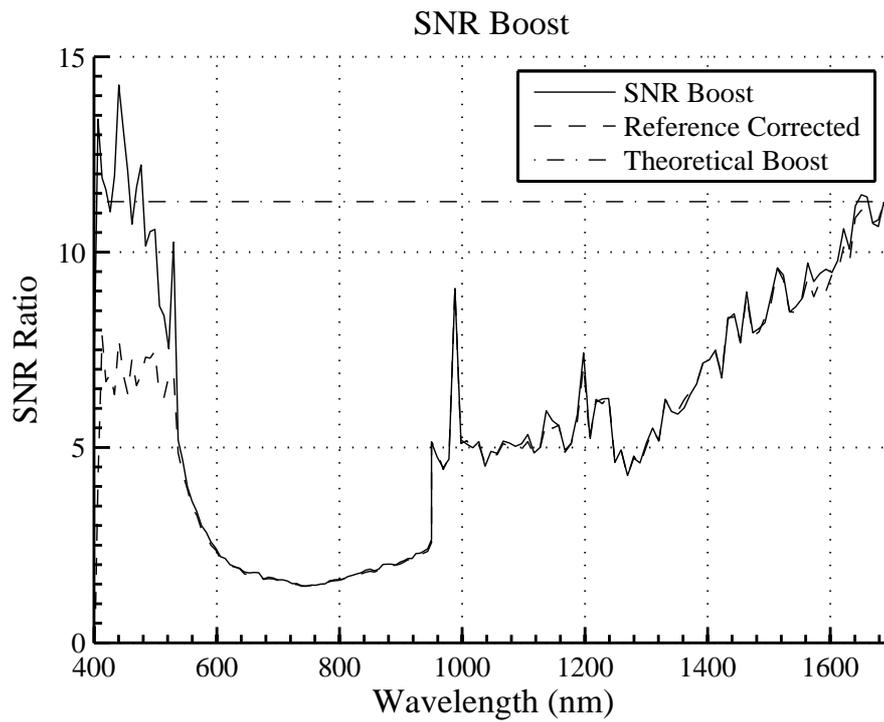


Figure 4.19: The SNR boost over pointwise imaging due to G-matrix encoding.

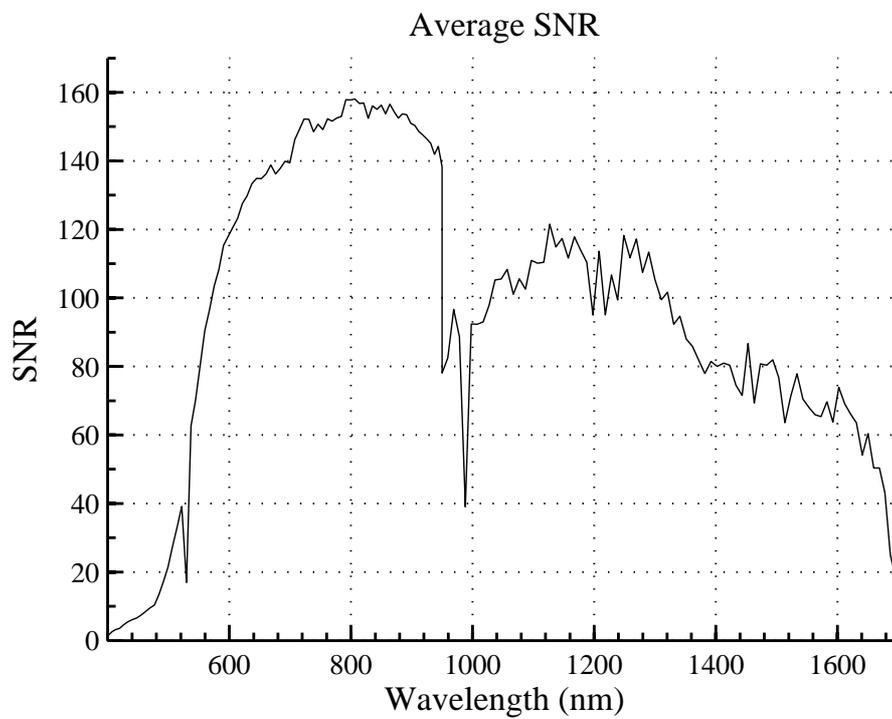


Figure 4.20: The average SNR due to H-matrix encoding.

spectra were reference corrected per spectrum before decoding. The SNR is greatest about 800 nm, has a sharp discontinuity near 950 nm and is very low in the visible domain. The G-matrix SNR has similar features.

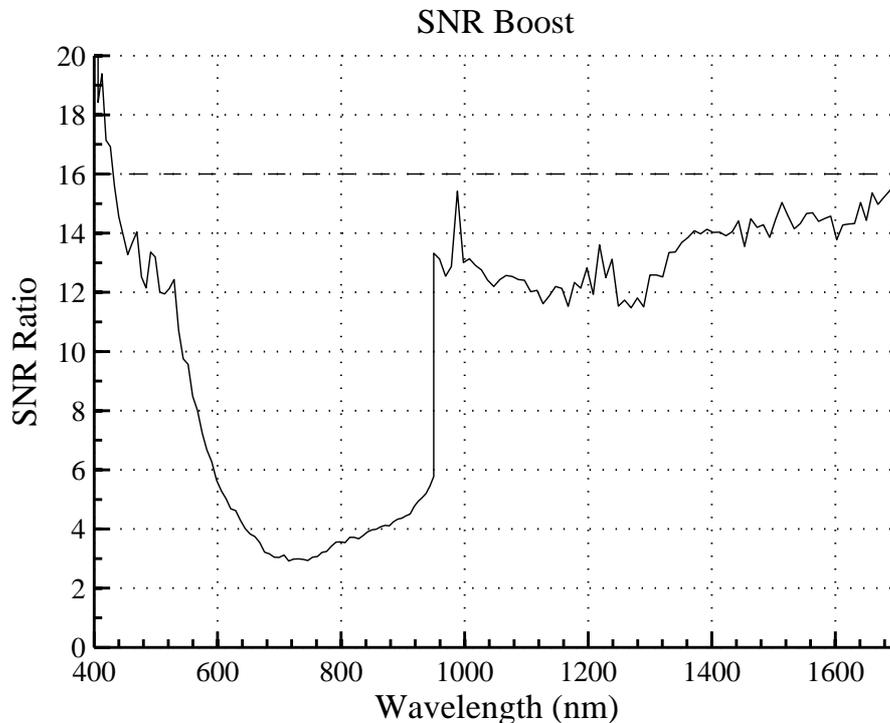


Figure 4.21: The SNR boost over pointwise imaging due to H-matrix encoding.

Figure 4.21 shows the SNR boost for H-matrix encoding over pointwise imaging with reference correction. The horizontal line shows the theoretical maximum boost of 16. The boost curve shows the same general shape as the G-matrix boost effect but is greater in value. The curve shows maximum boost at the highest and lowest wavelength ranges where the light intensity is least, shows minimum boost about 700–800 nm where the light intensity, hence photon noise, is greatest and has a sharp discontinuity near 950 nm at the change in the sensor material. Notably the greatest SNR boost is delivered in the low light level regions where it is most needed. The relative drop in SNR due to Poisson noise is not as pronounced for the H-matrix as for the G-matrix. For example between 1000–1200 nm the H-matrix boost is about 3/4 of the maximum possible, where for the G-matrix the SNR boost is about 1/2 the maximum possible in the same spectral region. Furthermore that for the H-matrix encoding we used a Spectralon white tile whereas for the G-matrix we used a white acrylic and cardboard object. Thus the H-matrix multiplexing was subjected to slightly more photon noise than the G-matrix (the reflectance ratio of the two objects is seen in Figure 4.14). The theoretical prediction that H-matrix multiplexing is better suited to situations where both additive and Poisson noise is present is confirmed.

Figure 4.22 shows the ratio of the SNR boosts from the H-matrix to G-matrix encod-

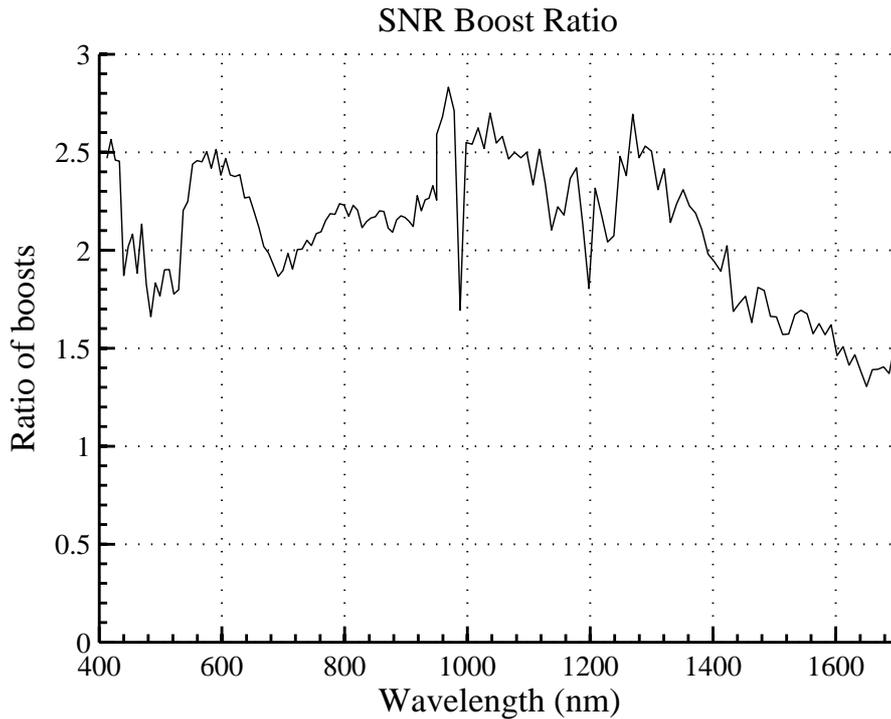


Figure 4.22: The ratio of the H-matrix SNR boost to the G-matrix SNR boost.

ing. For additive noise only the H-matrix was predicted to outperform the G-matrix by a factor of  $\sqrt{2}$ , as is seen in Figure 4.22 above 1600 nm. For Poisson noise the H-matrix is predicted to outperform the G-matrix by a factor of 2, which is the theoretical maximum of the ratio of the two boosts. From 400–1400 nm the SNR boost from H-matrix encoding is about twice that of the G-matrix, ranging up to around 2.5 times between 950–1150 nm, suggesting that Poisson noise is dominant in these regions.

The reason why the ratio of the H-matrix to G-matrix multiplexing boosts exceeded the theoretical prediction is not precisely known. It is possible however that other noise effects not explicitly discussed come into play. We can only conjecture about what these effects may be. For example if the photon noise of ambient light during the H-matrix acquisitions was greater than during the G-matrix then more encoding independent noise would be present, increasing the SNR boost. Also mechanical vibration in the system may have degraded the signal and reduced the SNR boost unpredictably.

### 4.3.3 Comparison of Compressed Sensing with Hadamard Imaging

Figure 4.23 shows the average reconstructed rows from the Hadamard technique of the Spectralon tile. There is a strong almost periodic pattern that arises from the concatenation of the rows. The low intensity (y axis) is due to normalisation by the highly intense reference beam.

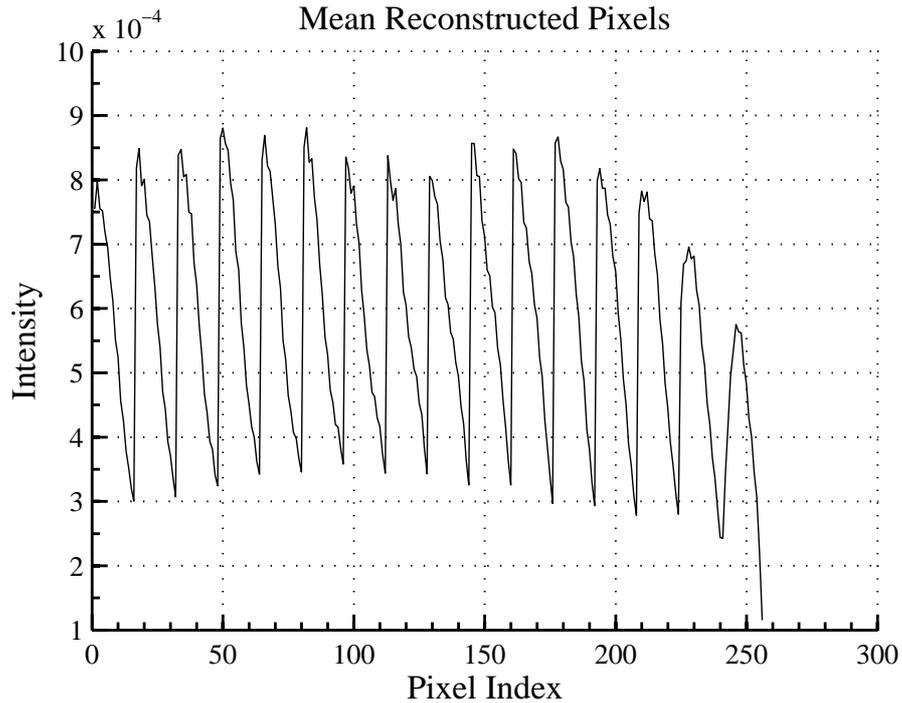


Figure 4.23: The average concatenated rows from Hadamard imaging of the Spectralon tile. This represents the best estimate of the true pixel values.

Figure 4.24 shows the SNRs for Hadamard, CS and pointwise imaging. Recall that the Hadamard and pointwise imaging use 256 measurements in accordance with the number of pixels. The CS reconstruction in Figure 4.24 was performed with 150 measurements. Interestingly the noise aware CS has not performed as well as the non-noise aware case. The SNRs in each case peak around 600–900 nm which corresponds to the peak output of the light source. Also there is a noticeable discontinuity near 950 nm that is due to the use of two types of sensor material in the spectrometer, one for the visible and very near infrared and the other for the near infrared. The Hadamard encoding outperforms the other techniques on our system in terms of SNR, particularly at the lower wavelengths where the photon noise is greater. In the near infrared region the non-noise aware CS approaches the Hadamard imaging.

Figure 4.25 shows the SNR boost of Hadamard and CS imaging over the pointwise imaging. The CS reconstruction was performed with 150 measurements. Again the Hadamard imaging outperforms CS on our system. Both methods have a SNR boost greater than one which means that they provide an advantage over pointwise imaging. In both Hadamard and CS there is a sharp drop in the SNR boost where the output of the light source and sensor response is greatest. This drop in SNR boost is therefore dependent on the relative level of photon and instrument noise.

Figure 4.26 shows the peak SNR of the CS against the number of measurements used

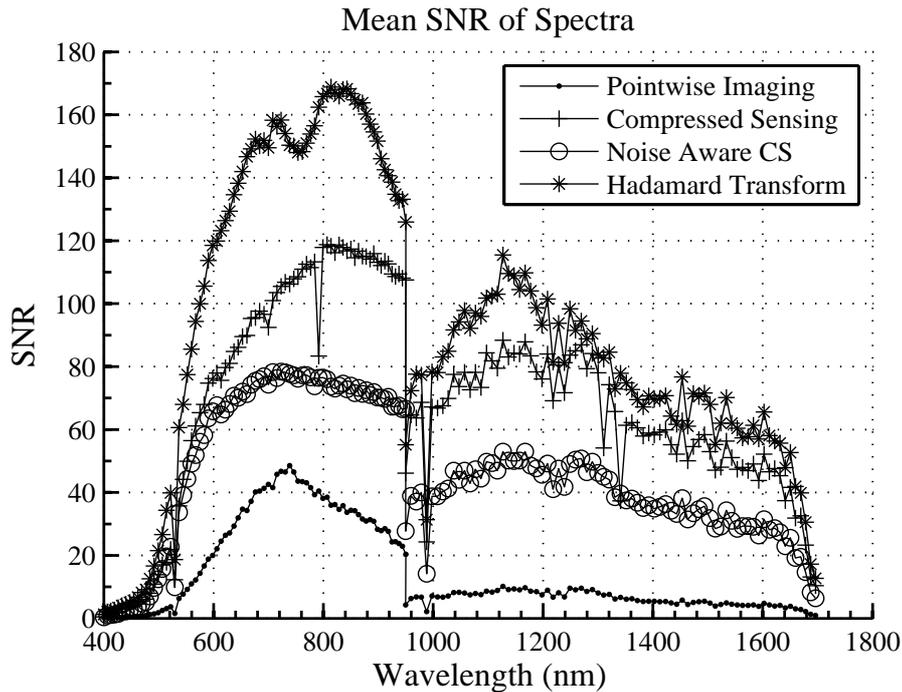


Figure 4.24: Signal to noise ratios for Hadamard, compressed sensing and pointwise imaging.

in the reconstruction. The non-noise aware reconstruction has an negative almost linear relationship with the number of measurements. CS seeks a sparse solution, so fewer measurements causes a solution with more entries near to zero. This in turn leaves less room for random fluctuation in the smaller coefficients. The shape of the noise aware CS SNR in Figure is similar to that of the theoretical SNR shown in Figure 3.1. In Figure 4.26 the experimental SNR was computed by comparing the reconstructed signal with the known true signal. Here the SNR was computed by examination of the fluctuation around the average reconstructed signal, so is more similar to the theoretical prediction, especially as the SNR does not drop at a low number of measurement. It is worthwhile to note that while the SNR, hence repeatability, from fewer measurement is higher than from more, it is likely to be repeatably inaccurate. The noise aware CS reconstruction achieves maximum peak SNR at 130 measurements, but does not reach the Hadamard peak SNR. Also the noise aware reconstruction is much closer in SNR to the pointwise imaging than to the Hadamard imaging.

Figure 4.27 shows the reconstruction error percentage against the number of measurements used. The non-noise aware reconstruction follows a typical shape in CS reconstruction for a non-sparse compressible signal. The error in reconstruction increased with the SNR, a trade off between accuracy and precision of measurement. At 256 measurements there is a slight upturn in error, presumably due to the reconstruction not taking account noise effects. The noise aware reconstruction is very inaccurate. Despite the peak SNR

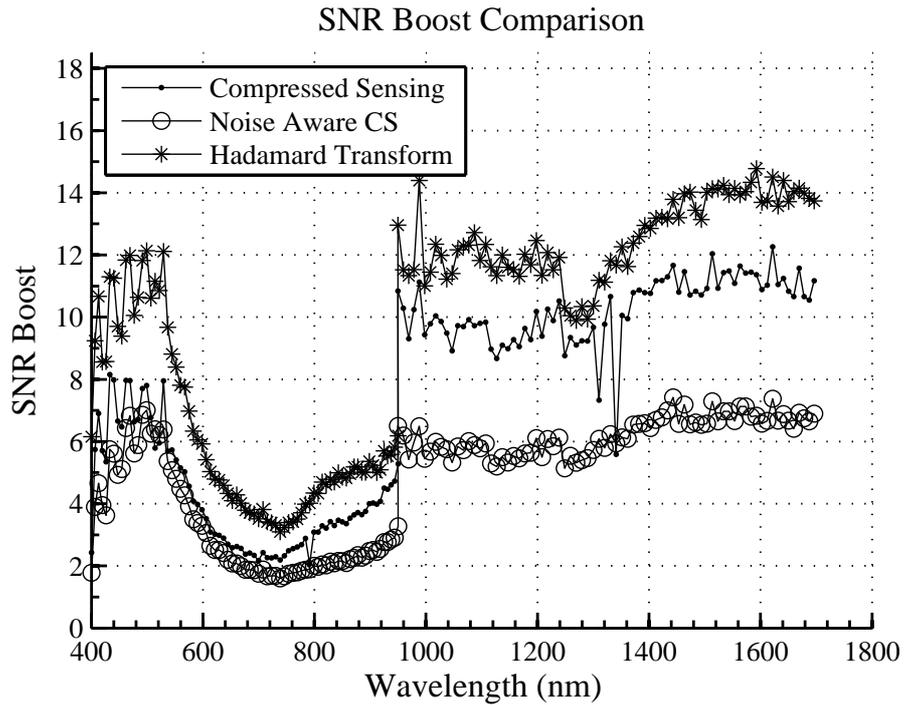


Figure 4.25: SNR boost comparison between Hadamard encoding and CS.

for noise aware reconstruction being greatest at 130 measurements the corresponding reconstruction error is greater than 60%. So while the noise aware CS reconstruction may be repeatable at 130 measurements, it is repeatably incorrect. Clearly the data acquired do not satisfy the model of noise aware reconstruction used by the basis pursuit algorithm.

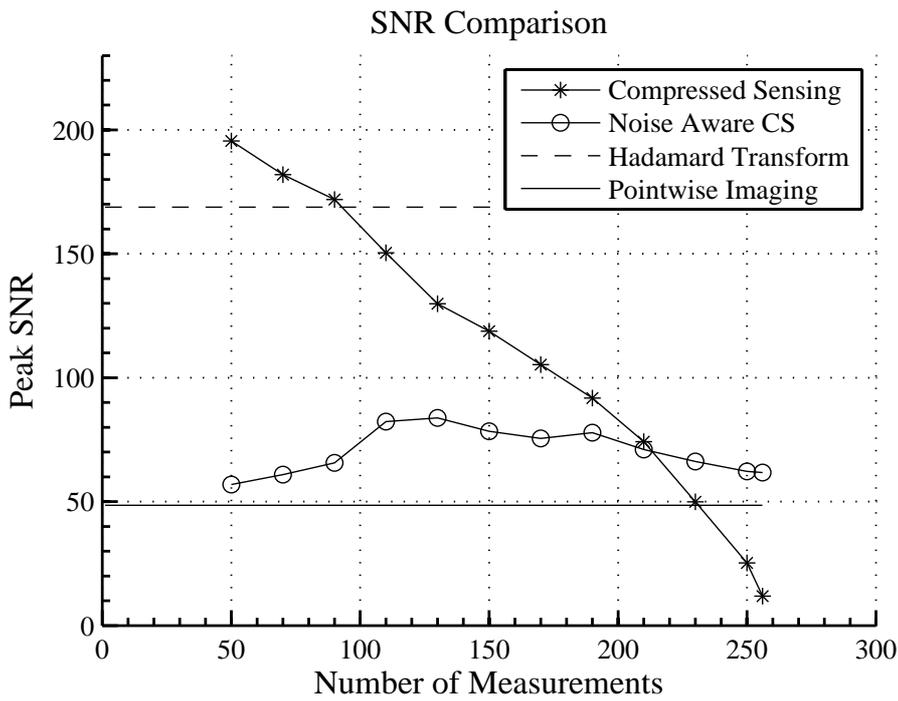


Figure 4.26: SNR against number of measurements for CS. The Hadamard and pointwise lines are shown purely for reference and do not imply Hadamard or pointwise imaging with fewer than 256 measurements.

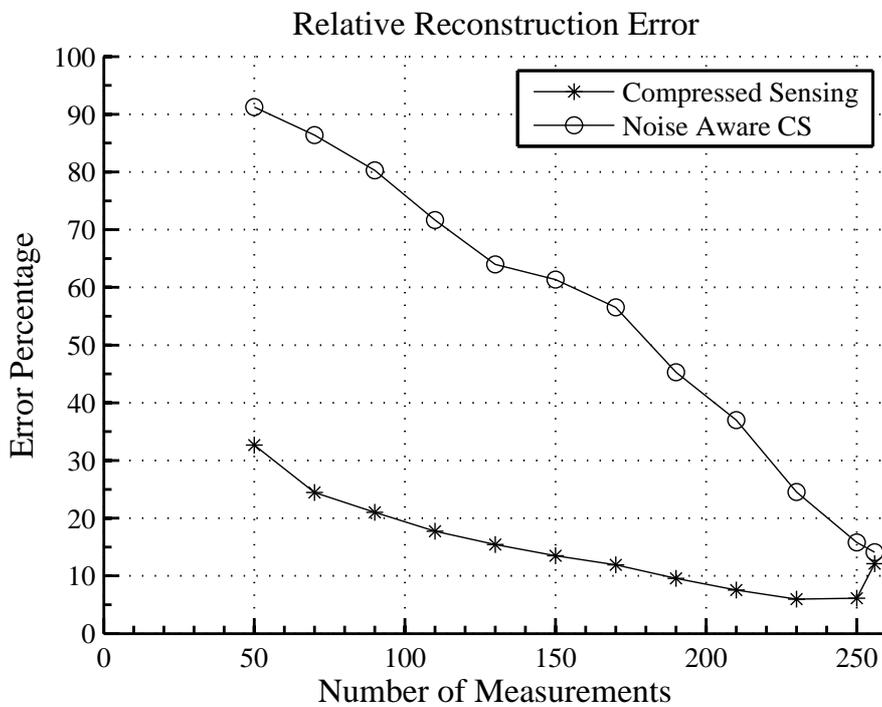


Figure 4.27: Relative error in CS reconstruction against the number of measurements.

## Chapter 5

# Volumetric Mapping of Beef

## *M. Longissimus Dorsi*

This chapter is concerned with an application of the hardware to prove the concept, that is, to show that the hyperspectral imaging system produces spectra that are of sufficient quality for use in practical applications. Ultimately it would be useful to have a general system for chemometric applications with spatial mapping. The primary application however is to map spectral variability in products and produce.

To demonstrate the applicability of the imaging technique we examine variability in beef *M. Longissimus Dorsi* (LD)<sup>1</sup>. The LD, also known as the strip loin, is a compound muscle that runs along each side of the spine from the 12th vertebrae to the tail end of the animal. The LD is the source of many important retail cuts such as the porterhouse steak, the ribeye steak and beef rib steak. NIR spectroscopy and analysis represents significant potential in nondestructive grading of meat. As such it is a topic of investigation, see section 2.7.

Ultimately an indication of quality features in the entire carcass from as few measurements as possible is desired. A grading system might return indicators of meat quality from possibly one measurement at a specific location on the animal. Meat grading by NIR is an ongoing project at AgResearch, with the objective of grading a whole carcass from a few measurements. Such lofty goals are not aimed for in this work. Here we examine the variability in the important LD.

When analysing spectra of meat to perform some estimation of the meat quality it is important to ensure that the spectra are of lean. If a given spectrum is of a part of the meat that has a high proportion of fat, connective tissue or even air, then an estimation of quality features from that spectrum might be unacceptably biased. For example it makes no sense to measure pH or tenderness of fat, so estimation of such features must not be performed

---

<sup>1</sup>A description is found at <http://bovine.unl.edu/bovine3D/eng/muscleIndex.jsp>, date accessed 4/3/2009. Identified as 'Longissimus'

on spectra of fat. Thus we examine classification of spectra into lean and nonlean groups. To verify classification colour imaging and image processing was used as the reference method. The classification is therefore against visible content.

Characterisation of the nature of the variation in the spectra down the length of the muscle was performed by variography. Variography estimates the correlation of any point to any other point a certain distance away, thus variography reveals features such as increasing or decreasing similarity and periodicity with distance. We also piggyback some basic quality assessment pilot studies of the meat. The pH down the length of the LD and between LDs, tenderness and blooming of the meat with NIR spectroscopy are examined. These extra quality features, in particular the pH and tenderness, inform about the variation in the samples examined. As described below in Section 5.1.1, steps are taken to prevent shortening of the meat so a broad range of tenderness and pH values are not expected.

## 5.1 Meat Collection and Instrumental Apparatus

This section outlines the data collection apparatus, collection scheme and describes the analysis procedures used. LDs were collected early post slaughter and stored at until rigor, on which spectra and colour images were taken. A steak from the head end of the muscle was subjected to tenderometry readings and a bloom curve over one hour after slicing was acquired. Down the length of the LD muscle hyperspectral and colour images were taken of successive slices. Image processing of the colour images was utilised to determine the visible content of each spectrum in each slice. Classification of each spectrum as lean and nonlean was performed using chemometric analysis. Variography was applied to the slice spectra per wavelength down the length of the muscle. Measurements of pH were taken on each slice and the pH correlated with the spectra.

### 5.1.1 Meat Collection

Fourteen beef *m. longissimus dorsi* with connecting muscles were harvested. The left and right sides of seven animals were taken, cut from between the 12th and 13th vertebrae and the tail. The breeds were four Angus cross and three Hereford cross. The tail muscle was removed from the LD then the LD was rolled in clingwrap to prevent shortening (Devine *et al.*, 1999) with one continuous piece that wrapped around the muscle five times. The wrapped LDs were stored at 15°C for 72 hours to allow onset of rigor.

### 5.1.2 Instrumental Configuration

The wrapped meat was placed in a stainless steel tube that was elevated above the equipment table. The data collection apparatus consisted of the hyperspectral imager, a Jai CV-M90 RGB colour machine vision camera (Copenhagen, Denmark) with a Fujinon CL11052 Closeup TV Lens (Saitama City, Japan), a Mettler Toledo 1140 pH meter (Greifensee, Switzerland) with a Mettler Toledo InLab 427 probe, and a MIRINZ meat tenderometer (MacFarlane and Marer, 1966).

Figure 5.1 shows a diagram of the imagers for the experimental setup. The hyperspectral image is comprised of the light pattern projector and the spectrometer, with meat in the sample holder completing the optical pathway of light from the source to detector. The colour camera and the light projector are trained on the meat face along very close to the optical axis' (causing specular reflection issues discussed below). Light from the projector is incident on the sample at approximately  $5^\circ$  to the meat face surface normal and the camera is slightly elevated above the projector and at normal angle to the meat face from the plane view. The spectrometer views the meat face at approximately  $60^\circ$  on the same side as the projector, so specular reflection from the normal surface on the sample is at  $65^\circ$  degrees to the spectrometer. Use of the tube for presentation of the meat provides consistent positioning with respect to the imagers. The hyperspectral imager light source, with all the mirrors on the DMA set to the on position, is also the illumination for the colour imager.

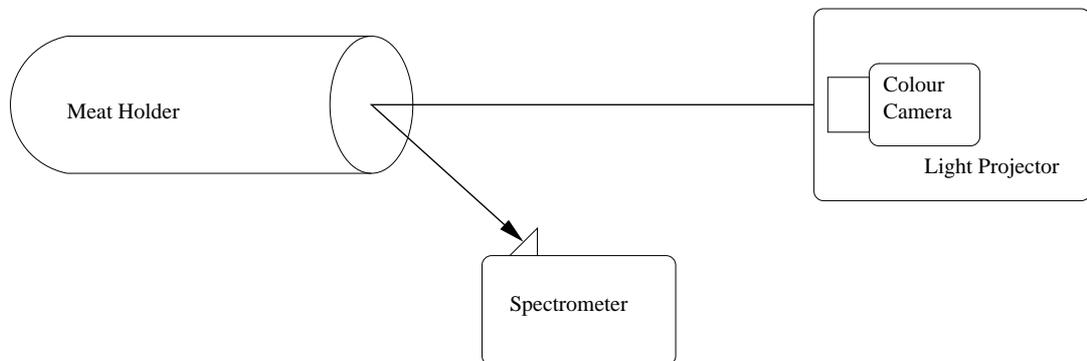


Figure 5.1: Diagram of the experimental setup.

### 5.1.3 Data Collection

During data collection the room temperature was set to  $16^\circ\text{C}$ , but ranged from  $15\text{--}17^\circ\text{C}$ . A first steak around 120mm thick, but at least 100mm thick, was taken from the head end, cooked by boiling and subjected to tenderometry readings. The meat was placed in a non-sealed plastic bag and immersed in  $100^\circ\text{C}$  water and monitored continuously

until the centre reached 75°C. Slices were cut 10 mm wide from which ten bites were cut 10 × 10 mm and about 30 mm long. The ten bites were sheared at right angles to the muscle fibre direction using a MIRINZ tenderometer (MacFarlane and Marer, 1966) to measure the shear force. The average shear force of the ten bites was taken as the shear force measurement.

Immediately after the initial slice the bloom readings began. A colour image, hyperspectral image rotation proceeded for one hour to sufficiently capture the breadth of colour change during blooming.

Twenty 13 mm slices were taken, proceeding down the length of the muscle to volumetrically map the LD. At least three colour images were taken of each slice immediately after it was cut. Fluid on the meat face caused some specularity in the colour images. Blotting with absorbent paper before each colour image reduced the specularity, nonetheless some specularity remained but the repeated blotting caused the specularity to appear in different locations in each image. After taking the colour images a 16 × 16 ( $N = 256$ ) Hadamard hyperspectral image was acquired. The slow spectral image acquisition time (approximately 2.5 minutes per image) and the time constraint that we sought to complete the mapping of two LD muscles per day are the reasons why a low spatial resolution was used. The next slice was cut and the pH measured on the meat just imaged. Five pH measurements of each slice were taken.

### Colour Image Processing

The colour image processing was similar to that of Streeter *et al.* (2006a) (which was in turn informed by earlier work (Streeter *et al.*, 2005)) but with some adaptations and improvements. Despite the blotting, some specularity appearing as small bright regions was present in all the colour images. The blotting used to reduce the specularity also ensured that most specularities occurred in different locations in each image. The minimum over the set of colour images at each pixel for each meat face was calculated, producing one colour image per slice that was cleaner than the raw images.

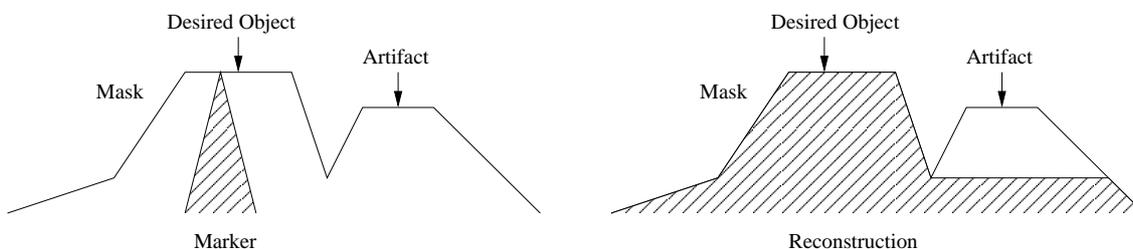


Figure 5.2: Illustration of the action of the morphological reconstruction. On the left is the mask and the marker and on the right is the result after reconstruction.

After the minimum operation further cleaning of the image was performed with a

morphological reconstruction to each colour plane (Vincent, 1993). The morphological reconstruction operation requires two images, a marker and a mask where each pixel in the marker is less than or equal to the mask pixel at the same location. The minimum image was used as the mask and a morphological opening with a  $3 \times 3$  structuring element of the minimum image as the marker. The marker image indicates where to begin and the mask determines the maximum pixel value the reconstruction can attain. Beginning from the pixel values in the marker image, morphological reconstruction spreads out laterally until a pixel in the mask image indicates to stop. Figure 5.2 illustrates the action of the morphological reconstruction. The mask in the illustration contains two objects, one is desired and the other is an artefact. The marker indicates which object is desired and the reconstruction fills out this object. The artefact is suppressed because the reconstruction cannot return values greater than the boundary between the artefact and the desired object.

The region just inside the meat holder tube was manually located and masked in the image to remove the metal tube and outside region. The image was transformed into CIE  $L^*a^*b^*$  colour space. A threshold value was computed for, and used on, the  $a^*$  plane using Otsu's method (Otsu, 1979). The result of the threshold formed an estimate of the lean region. Fat was identified as bright areas in the green plane of the RGB image. The green plane was selected because red objects were minimised. The blue plane was not useful as the images were taken directly after slicing, so had a purplish appearance due to deoxy myoglobin (blooming had not occurred). The fat surrounding the lean was identified using a different threshold computed again using Otsu's method on the green plane. The resultant fat estimate was refined by excluding the lean region and the region outside of the meat. Excluding the lean region also excluded marbling. Marbling appeared as bright objects inside the lean region. These marbling objects were identified as statistical outliers in the green plane assuming normal statistics. Any pixel in the lean region with intensity at least three standard deviations greater than the mean was labelled as marbling. The final fat binary image was the combination of the outer fat and marbling. Streeter *et al.* (2006a) used a pattern recognition classifier method to improve the specificity of marbling object detection. Here the colour image acquisition was sufficiently improved so that the segmentation was considered specific enough to not warrant a further classification.

Figure 5.3 shows two (of three) replicate images of the same slice, the result after taking the minimum and morphological reconstruction and the colour segmentation. Small bright specularities are in both the colour images on the top row, but appear in different locations on the face of the slice due to the blotting of the meat face with absorbent paper. Around the edge, especially on the top left there is a bright reflection due to the plastic wrapping used on the meat. The plastic wrapping causes bright areas that confuse the fat detection scheme, as seen in the segmentation image on the top left. A morphological

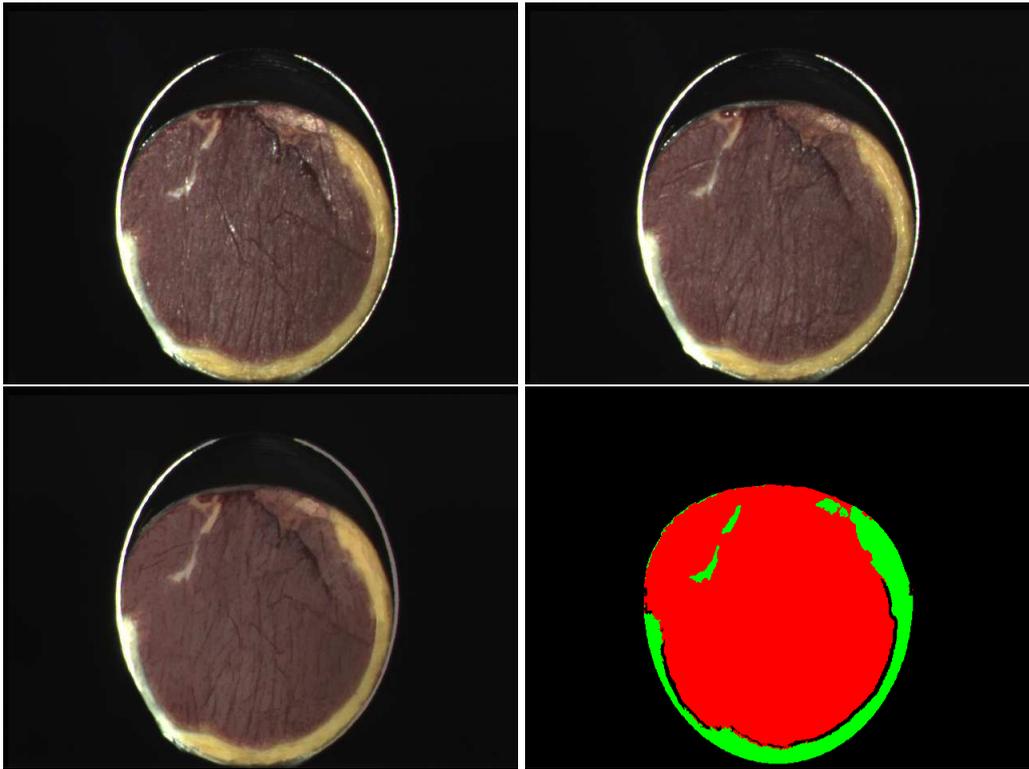


Figure 5.3: Images of animal 5, left hand side, slice eight. The top row shows two different colour images of the same slice with blotching between captures. The bottom row, left shows the preprocessed image formed from the images of the same slice. Bottom row, right, shows the segmentation where red is the pixels classified as lean and green as fat.

opening was considered to remove the edge region on the final segmentation to remove false fat objects due to the plastic wrap, but was dismissed because opening changes the morphology of the object which is not desired.

The colour image processing result was registered to the hyperspectral images to enable estimation of the visible content of the meat for each spectrum. To perform the registration, two complement checkerboard patterns with squares corresponding to the hyperspectral image pixel locations were projected on to a flat white object in the place of the imaged meat surface. Images were taken of the checkerboard patterns, were processed and the locations of each square registered. In this way the areas on each colour image that corresponded with each pixel in the hyperspectral images were identified. The reference visible content proportion for the hyperspectral image pixels was calculated. Spectral pixels were identified as lean if there was more lean than any other visible constituent.

We have the opportunity here to assess not only the quality of the primary analysis method (classification of NIR spectra) but also the reference method. The colour image processing is essentially the automation of class assignment that otherwise would have been performed manually. Manual rechecking of the colour image processing result allows us to obtain an estimate of the error in the reference method. Ten slices were

selected at random from the dataset. The original images and the colour segmentation image were displayed on screen. Regions in the image where incorrect segmentation had occurred were manually ‘clicked’ and the pixel locations recorded, with even and moderately dense coverage of the incorrect regions aimed for. Spectral pixels with at least one identified error location were labelled as having error. Having some error does not necessarily mean that the spectral pixel was incorrectly labelled, it means that mislabelling could have occurred, so the method of estimation of reference error employed indicates the upper limit of the error level. The ratio of spectral pixels with error to total number of spectral pixels was taken as an estimate of the upper limit of error in the reference method. For each image the percentage of spectral pixels with error in the reference was computed, and the mean and standard deviation of the error percentages were taken over the ten images.

#### 5.1.4 Data Summary

Table 5.1: Enumeration of the breeds from which the muscles were taken.

Animal/Day	Breed
1	Angus cross
2	Hereford cross
3	Hereford cross
4	Angus cross
5	Angus cross
6	Hereford cross
7	Angus cross

The LDs were taken from steer carcasses, shot with captive bolt and electrically stimulated and harvested from breeds on an as available basis. Table 5.1 lists the breeds per animal. The two LDs for each animal were used for each day of data collection, so a reference to day three, for example, corresponds to animal three.

The hyperspectral images of the volumetric mapping consist of twenty  $16 \times 16$  pixel images per muscle over the fourteen muscles corresponding to seven sets, that is, one set per animal. In total this gave a raw data set of 71680 slice spectra. Spectral pixels of the air around the meat face were considered outliers and were easily identified by

Table 5.2: Breakdown of the number of spectra per animal obtained after the removal of air spectra.

Animal	1	2	3	4	5	6	7
No. of Spectra	8347	7952	6626	6348	8004	8556	6380
No. of Lean Spectra	6647	6449	5495	5240	6600	6725	5123

Table 5.3: Summary of the reference data. The variables  $a_0^*$  down are defined in Section 5.4.

Variable Name	n	Mean $\pm$ Standard Deviation	Minimum	Maximum
Tenderometry(kgF)	14	5.92 $\pm$ 1.10	4.26	7.75
pH	12	5.37 $\pm$ 0.15	5.08	5.75
$a_0^*$	12	17.6 $\pm$ 2.47	14.7	22.9
$a_f^*$	12	20.4 $\pm$ 2.75	17.0	25.9
$k_a$	12	0.172 $\pm$ 0.075	0.0684	0.322
$b_0^*$	12	6.71 $\pm$ 2.02	3.01	11.4
$b_f^*$	12	10.6 $\pm$ 2.09	7.60	14.8
$k_b$	12	0.225 $\pm$ 0.103	0.0854	0.440

simple thresholding and removed leaving 52213 spectra. Table 5.2 gives a breakdown of the usable nonair slice spectra for each muscle. The number of spectra per animal is determined by the size of the muscle. Bloom spectra are a different set of data and are not included in the counts listed in Table 5.2. The maximum error in the visible content reference was estimated to be  $3.7 \pm 1.6\%$  to two significant figures.

Table 5.3 summarises the statistics of the reference data. In each case there is one sample per muscle. For the pH twenty measurements per muscle were taken, but as detailed below, the intramuscular spread in pH was small compared to the variation between muscles. The parameters for the bloom represent one bloom curve per muscle.

## 5.2 Lean Pixel Classification

Classification of the hyperspectral pixels as lean dominant was investigated. The reference classification was designated lean dominant if there was more lean in that pixel than any other content. ECVA (see Section 2.5.3) was employed to build a linear classification model for the two class problem. The power of classification was assessed using ROC (see Section 2.6.3) analysis. We compute the ROC AUC using the trapezoidal rule and use the SE (equation 3.102) of the Wilcoxon to estimate the SE of the AUC. This work was informed by earlier investigations into visible content calibration Streeter *et al.* (2006b), specifically that study considered calibration of visible fat content by NIR.

The data were divided into three groups. The first group included all data except those from two animals. The second and third groups were comprised of the data of one animal only, each group being one animal not included in the first group. The second and third group were simply chosen as data from the last two animals acquired, however the choice was supported by the fact that a different breed was represented in each of the last two animals. The classification was analysed in three stages. The first stage was cross validation per-animal on the first group to probe if classification was possible. Per-animal

cross validation was used to maintain as much independence between groups as possible. The second stage was testing. Here ECVA models were built from the first group using a range of number of PLS factors, with the models tested on the second group. From the second stage the number of factors to use in the model was chosen. The third and final stage was validation. A model was constructed using the data from both group one and two together as one set, with the number of factors as determined in the second stage. The model was then used to classify the data from the third group and the AUC computed.

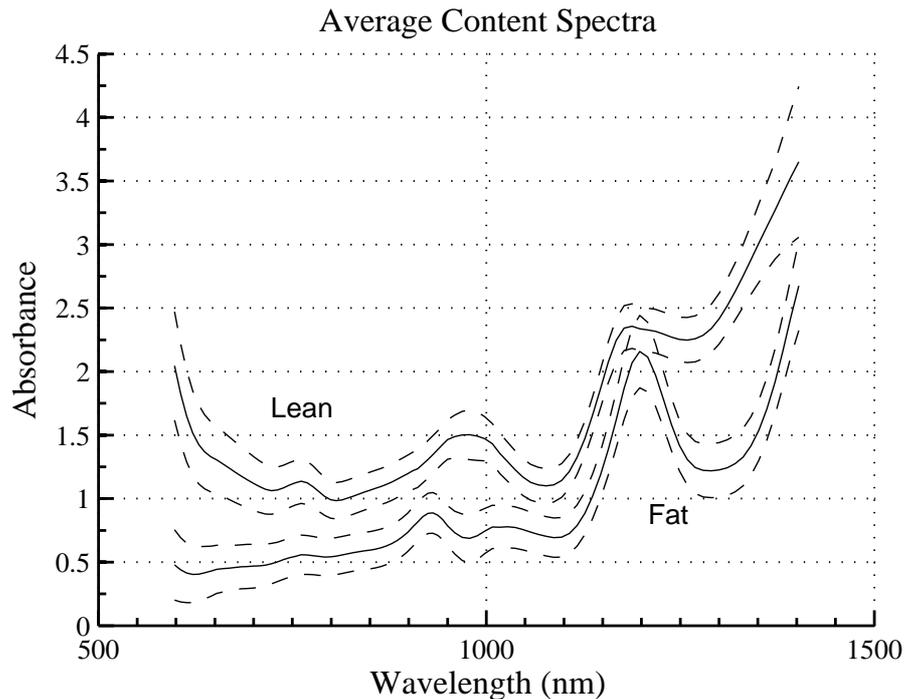


Figure 5.4: Average spectra from the slice images for days 2-6. The average lean and fat are shown as solid lines and the one standard deviation intervals are dashed lines. After removal of air spectra the fat is the dominant nonlean constituent.

The colour images from the first day's data acquisition were problematic and difficult to segment, so the last six days were used for pixel classification. The data from the six animals formed six subsets. Cross validation was used per-animal on the first four sets to probe the data for classification power. The fifth animal's data was used to test the model. The number of factors was chosen at the point where no increase in AUC (accounting for SE) can be obtained by adding more factors. To validate the model the spectra from the sixth animal was classified and the AUC computed.

Figure 5.4 shows the average lean and fat spectra with one standard deviation intervals, computed on spectra identified as completely lean and completely fat by the visible content reference. There are obvious differences in the shape with the most notable being that the lean has higher overall absorbance. The lean absorbance rises dramatically below 650 nm and at higher wavelengths towards the water absorption peak at 1500 nm. Around

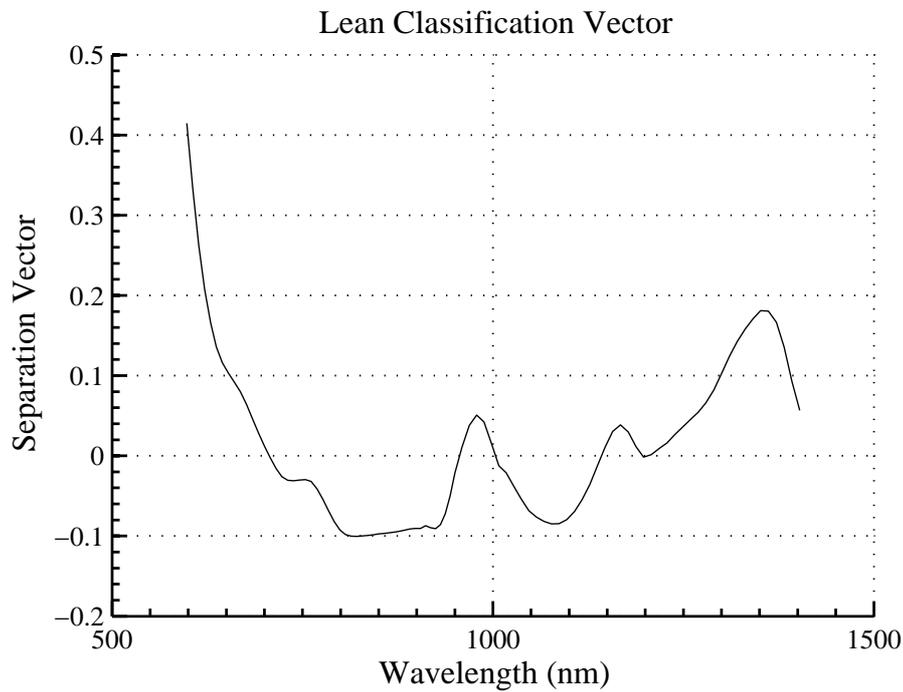


Figure 5.5: The classification vector. Regions greater than zero are indicative of the location of spectral features that distinguish the lean group from the nonlean group (the nonlean group is predominantly fat).

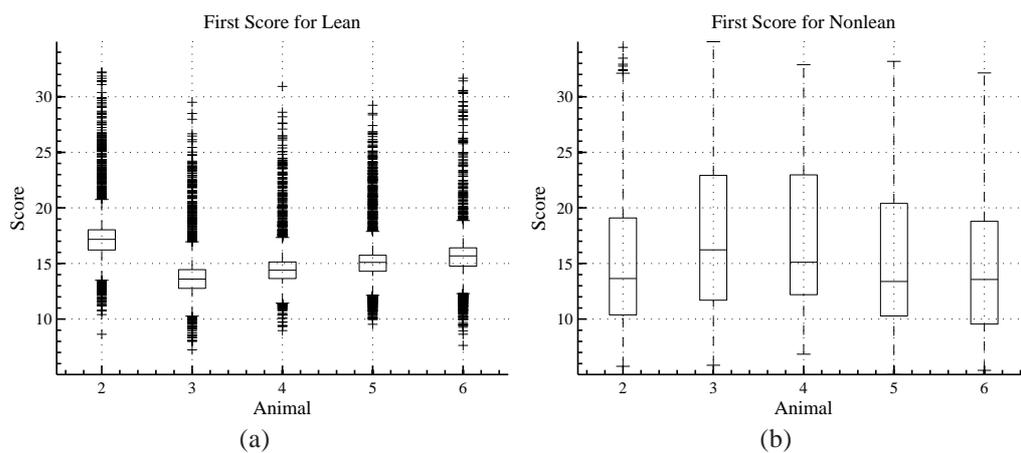


Figure 5.6: Box and whisker plot of the first score per animal, a) the lean group and b) the nonlean group.

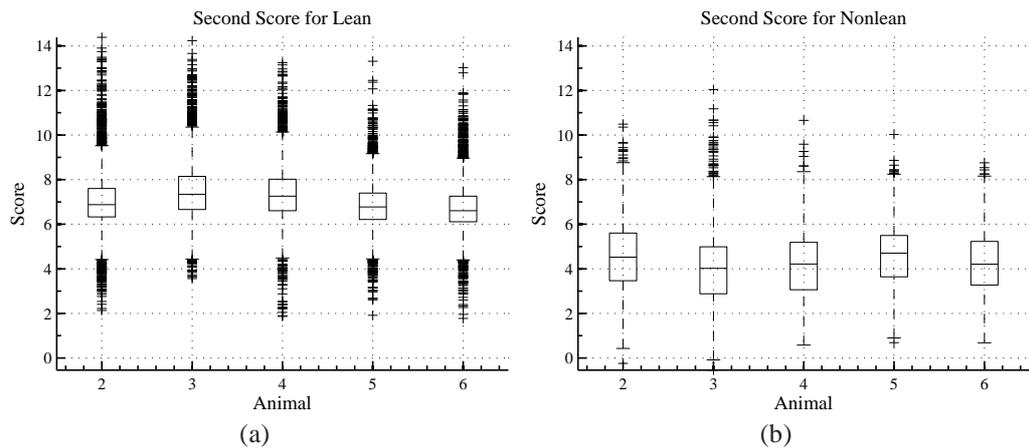


Figure 5.7: Box and whisker plot of the second score per animal, a) the lean group and b) the nonlean group.

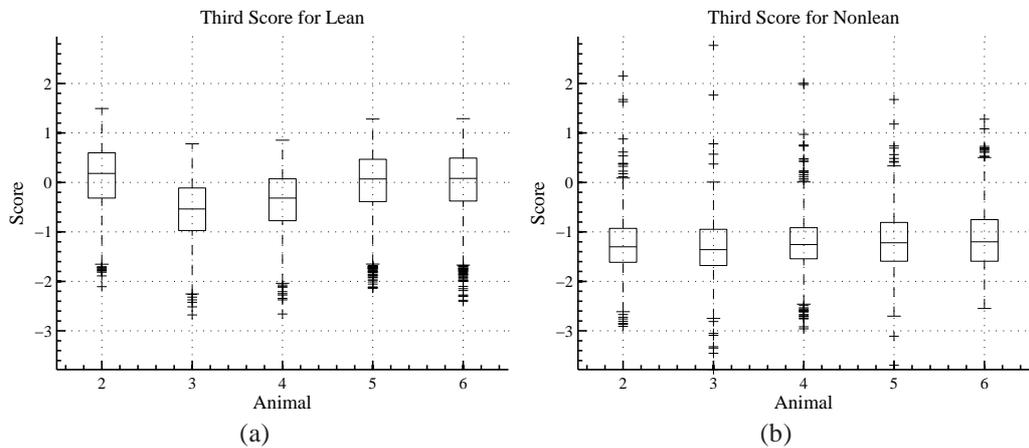


Figure 5.8: Box and whisker plot of the third score per animal, a) the lean group and b) the nonlean group.

1000 nm the lean has one peak where the fat has two peaks on either side. At 1200 nm there is an absorption peak due to the combination of the water and hydrocarbon peaks in both the lean and the fat. The lean spectra, being more influenced by the water, has its peak slightly to the left of the fat peak and does not decrease around 1300 nm like the fat spectra do.

Figure 5.5 shows the classification vector built from the data of five animals data during the validation stage. Notable features include the large weighting given to the rise in lean absorption at shorter wavelengths; the small bump at 750 nm corresponding with the small water peak in the lean; the peak at 950–1000 nm that sharply drops off below zero either side corresponding with the water feature in the lean and hydrocarbon features in the fat spectra at the same wavelength region; the peak just below 1200 nm and the large positive region above 1200 nm that peaks at 1350 nm corresponding to where the

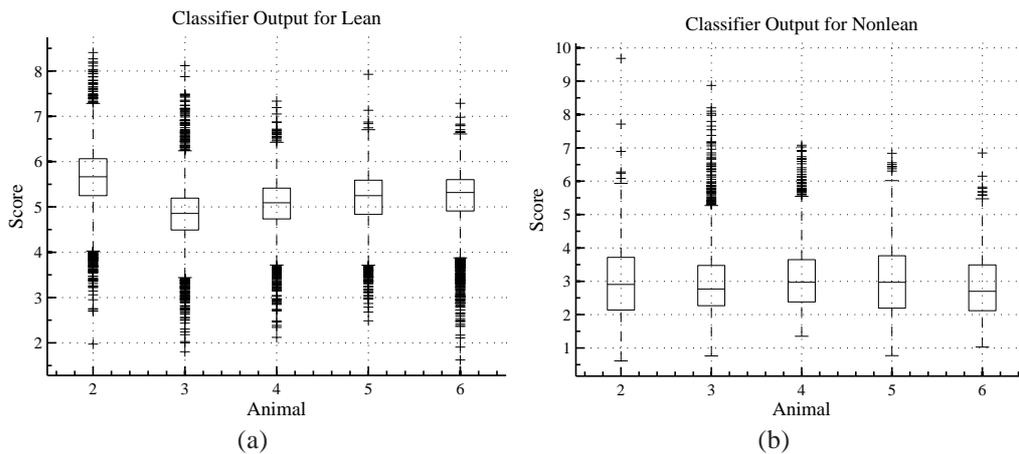


Figure 5.9: Box and whisker plot of the overall classifier output per animal, a) the lean group and b) the nonlean group. This is the result of the classification vector on the data it was computed on so is likely to be favourably biased.

separation between the two groups of spectra is large due to the edge of the large water peak in the lean and fat, but is more dominant in the lean than the fat. Qualitatively we see that the ECVA classifier has sensibly utilised the major features of difference between the lean and nonlean groups.

Figures 5.6, 5.7 and 5.8 show box plots respectively of the first second and third score of the training data loadings per animal per group. The factors shown are due to the ECVA computed at the training for the validation stage. The score for none of the animals is obviously different to the rest, barring perhaps animal two for the first loading on the lean group, but animal two is not greatly different to the rest. The lean and nonlean scores for animal three and four are not separated as well as the second, fifth and sixth in the first score. Figure 5.4 shows that gross absorbance is a strong discriminating feature between the lean and fat. The first score is heavily dependent on the difference in the average absorbance in the lean and fat groups. The variation in the first score seen in Figure 5.6 might lead one to conclude that this gross effect should be removed by preprocessing, however this gross absorbance difference is too important a spectral feature to be thrown away by using a preprocessing step such as the SNV transform. One could alternatively use MSC or EMSC (Section 2.5.2) so that the gross effect is normalised per animal while the gross difference between lean and fat is retained. Unfortunately such preprocessing requires prior knowledge of the animal which is not necessarily available when analysing a spectrum from a new animal at random. Still EMSC could have been used with an average spectrum over all the animals instead of per-animal, but was not considered necessary. Ultimately gross absorbance is an important feature but is not necessarily the strongest discriminating feature in all the natural groups (animals) present.

The first and second scores have a large number of points that look like outliers. These

supposed outliers do not sit at a large distance from any other measurements and are not due to grossly erroneous measurements or true outliers (such measurements were removed earlier with the air spectra). Furthermore any new measurement may reside in the ‘outlier’ regions so it is appropriate to include them in the calibration data set. Figure 5.9 shows the classifier output before thresholding per animal on the training data. The separation between the lean and nonlean groups is apparent, although the plots do show application of the classifier on the same data the classifier was computed on, so are likely to be favourably biased.

#Factors	Cross Validation	Testing
1	0.6824(0.0037)	0.6983(0.0063)
2	0.9048(0.0018)	0.9070(0.0031)
3	0.9614(0.0010)	0.9755(0.0014)
4	0.8923(0.0019)	0.9779(0.0014)
5	0.9702(0.0009)	0.9817(0.0012)

Table 5.4: AUC(SE) for pixel classification.

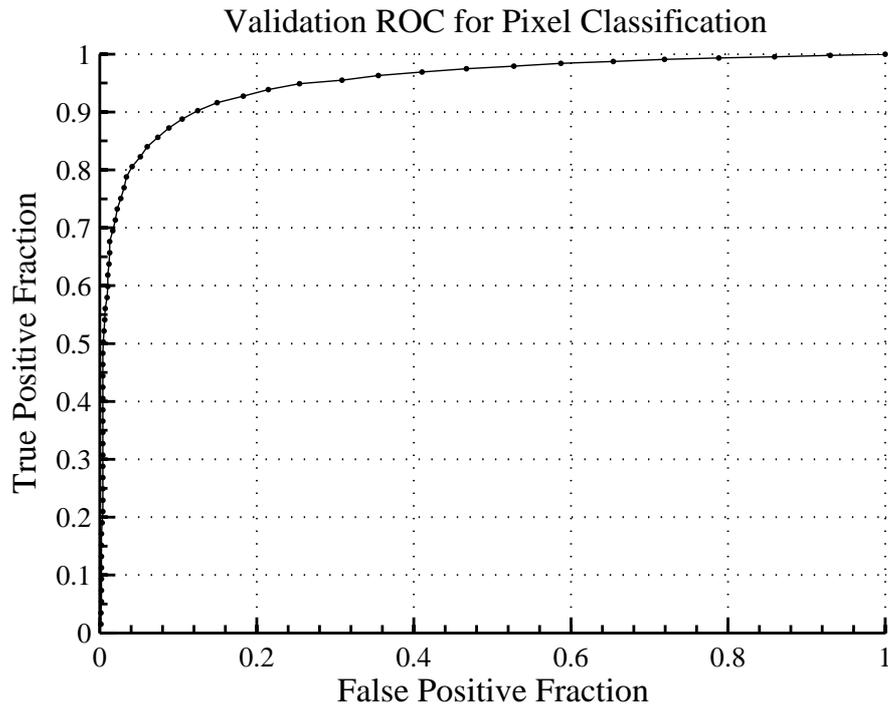


Figure 5.10: ROC curve for the validation data.

The AUC and associated SE for the cross validation and testing are shown in Table 5.4. In every case the AUC via the trapezoidal rule and the Wilcoxon were computed and found to be the same. In the cross validation the AUC increases from one to three factors and then drops at the fourth. The AUC(SE) for the third factor is 0.9614(0.0010) indicating

strong classification power. The test AUC increases until the third factor, beyond which the increase is small and the intervals indicated by the SE overlap, thus we select a three component ECVA model.

Figure 5.10 shows the validation ROC for hyperspectral pixel classification as visible lean. The validation AUC(SE) is 0.9515(0.0025). The ROC curve tracks close to zero false positives until  $TPF = 0.64$  where FPF starts to increase dramatically. The curve does not reach  $TPF = 1$  until  $FPF = 1$ , indicating that the spread of the true class (lean) reaches to the far side of the false group, that is, the minimum classifier result for the lean pixels is as small as the minimum result for the nonlean pixels. There are many more lean spectra than nonlean spectra, as evidenced by the large lean area compared to the fat area seen in Figure 5.3, so the spread of the lean classifier result is enough to reach the far side of the nonlean group. One might conclude that a Bayesian like approach is appropriate, in which the prior knowledge that there is more lean than nonlean is incorporated to bias the classifier towards classification of spectra as lean, increasing sensitivity. However the objective of this exercise is to identify specifically the lean spectra with a minimum of false positives so that analyses are guaranteed to be performed on lean spectra only. Biasing the classifier to classify spectra as lean runs the risk of reducing specificity.

Now we examine what the true value or the AUC might be taking into account the error in the reference. From Section 5.1.4 the maximum error in the reference is  $3.7 \pm 1.6\%$  and in the validation set there were 6380 spectra, of which 5123 were labelled lean by the reference method and 1257 were labelled not lean. We need to assess whether Equation 3.97 or (preferably, due to the simpler form) Equation 3.98 is more appropriate. We recompute the Wilcoxon on the classified validation data (before thresholding) one thousand times, each time using a subset of 1257 lean spectra, so that the lean and nonlean groups are balanced. The mean and standard deviation of the resampled Wilcoxon over the thousand random resamplings was 0.9514 and 0.0036, so the AUC figure above is within one standard deviation of resampled Wilcoxon. Because the figures are so close we approximate the error effect as a straight line and use equation 3.98. We highlight that if the error in the reference was much larger than it was in this case then the difference in the Wilcoxon may have been larger and the straight line approximation would not be appropriate, but in this case we can not conclude from the data that the straight line approximation is different from the nonlinear change in AUC due to error in the reference. Substitution of the data into the rearranged equation 3.98, namely equation 3.101, gives a maximum possible AUC(SE) of 0.98(0.02). So, concluding, the AUC of classification of the validation data is at least 0.94 (no error in the reference) and is at most 1 (error in the reference plus the SE).

Studies conducted by others focussed on fat content in meat, where we focus on identification of lean spectra. For example González-Martín *et al.* (2003) and Sierra *et al.*

(2008) examined the detailed content of different types of fatty acids as identified by chemical testing. We consider lean and fat in general as identified by visible content. However like them we see that NIR spectroscopy is effective for identification of the major constituents of meat, albeit we do not examine the detail that they do. Qiao *et al.* (2007a) performed marbling assessment using spectroscopy in the range 400–1000 nm, a band that overlaps ours but encompasses the visible domain which we have not been able to use here. They had much higher spatial resolution than what we have and were able to use features computed over the spatial dimension, whereas we performed classification using spectral information only. Regardless, classification of lean spectra is successful in this trial.

### 5.3 Variography

Variation in the spectra was examined down the length of the muscles by variography. The lean spectra were extracted from the data set using the derived classifier. The average spectrum for each slice was taken and the variogram of each muscle at each wavelength was computed. The average variogram for each wavelength was computed over the muscles and notable features of interest observed.

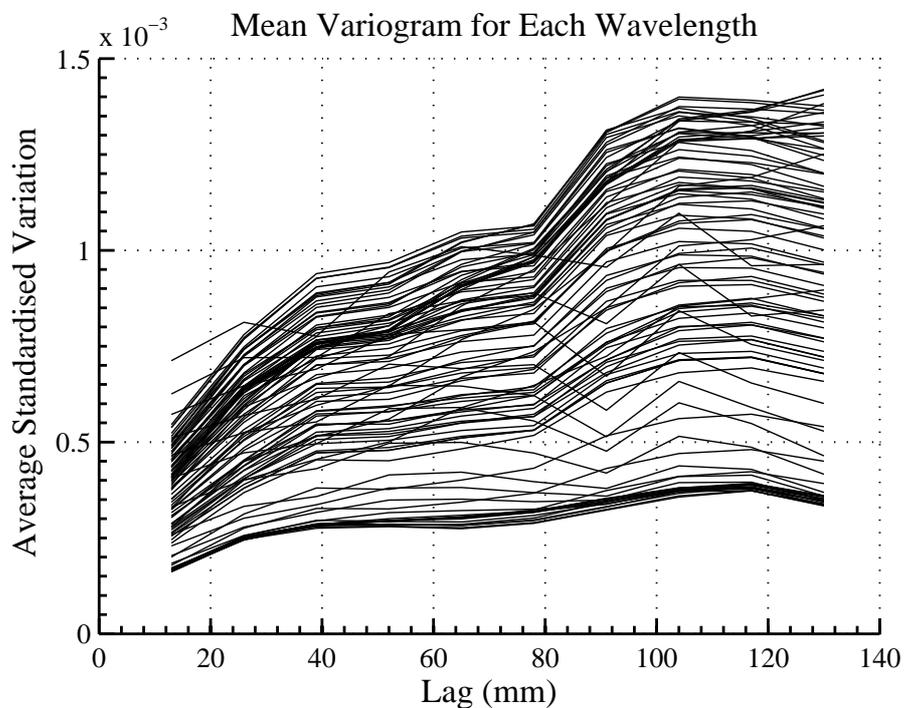


Figure 5.11: The average variogram for each wavelength down the length of the LDs. Each line represents one wavelength. The variograms for all the wavelengths are shown together in this manner so that the general shape of the variation against lag down the muscle can be seen.

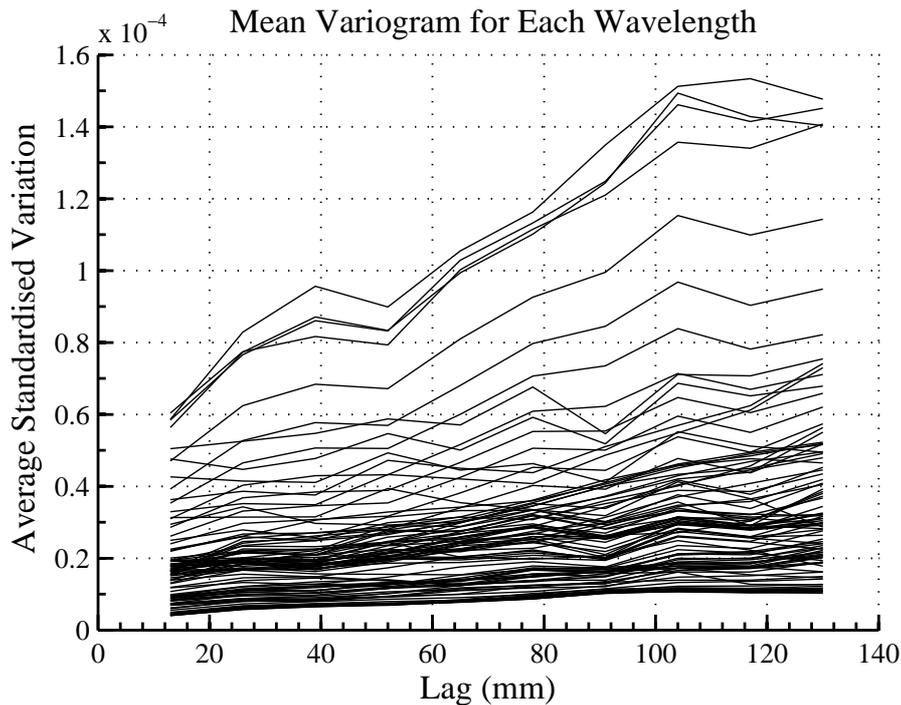


Figure 5.12: The average variogram for each wavelength down the length of the LDs after EMSC. The effect due to water and the average lean spectrum was removed before computation of the variograms.

The variography was performed down the length of the twelve muscles for days 2 to 7. The lean pixel classifier was applied to the spectra to segment the lean spectra and the average lean spectrum was taken for each slice. The variogram was taken for each muscle at each wavelength and the variograms averaged over the muscles resulting in one average variogram for each wavelength.

Figure 5.11 shows the computed average variograms. There are two obvious dominant effects: first the correlation between positions increases with distance down the length of the muscle and second there is a periodic effect in the variograms with period about 65 mm. Furthermore the trough in the periodic effect occurs at about 65 mm lag. The vertebrae in the animal are also about 65 mm in length. When the muscle is cut from the carcass an alternating pattern is formed of bare lean and lean covered with intact connective tissue, which has periodicity due to the the vertebrae, that is, about 65 mm.

There are two possible reasons why the vertebrae might cause a periodic effect down the muscle. The first possible reason is after the beast is slaughtered, and while it is butchered and dressed, the halves of the carcass are hung by the Achilles tendon. Mechanical stress from the weight on the carcass pulls down on the muscles, with some counter force at the connection between the vertebrae and the LD. The result is possible micro changes in the fibres of the muscle that cause a scattering effect on the spectra. The second possible reason is the connective tissue sheath around the LD causes moisture to

be retained in the lean, but where the muscle is cut from the carcass the connective tissue is removed from around the lean. Thus one might expect a periodic effect due to periodic variation of moisture in the meat. In section 5.2 we observed a number of features in the spectra that are due to water to varying degrees. These water dependent features proved useful in the discrimination between lean and nonlean groups (recall the nonlean group was predominantly fat). Again water appears to be a major source of variation in the spectra.

EMSC was performed on the average lean spectra and the variography repeated. Figure 5.12 shows the result. A representative water spectrum and the mean of all the slice spectra were used as the reference spectra in the EMSC. The water spectrum was acquired with a KES 200 series spectrometer by transmission of light through 1 mm of water at 20°C and reflected off a Fluorilon (Avian Technologies, Wilmington, OH, USA) white reference tile. The temperature of the water at spectral acquisition was chosen as the closest available in our database to the ambient temperature at data acquisition for the meat (15–17°C). Polynomial order zero was used in the EMSC. A polynomial order 2 was also tried but yielded no difference in the variograms. Two major differences that occur in the variograms after EMSC are the reduction of gross variation by a factor of ten and the removal of the 65 mm periodicity. There is still some shape to the variogram but nothing obviously consistent over the wavelengths. It is difficult to say whether the normalisation against the mean spectrum or removal of the water spectrum effect is predominant in removing the periodicity. The mean lean spectrum is greatly influenced by water so the mean effect is also highly correlated to the water effect. Regardless it is clear that the EMSC has removed the periodicity which arises either due to the water effect or some other effect that correlates strongly with the water. In future studies of spectra taken of meat not acquired in situ, particularly of the LD muscle, both the effect due to connective tissue removal from the exterior of a muscle, and that this effect is likely due to water seepage, should be born in mind.

To our knowledge this is the first study to examine spectral variation within the LD using variography. Hansen *et al.* (2004) and Janz *et al.* (2006) showed that variation in quality features occurs within other muscles. Variation within the LD is not surprising given that it is a relatively large compound muscle, so it is reasonable to expect variation to exist pre-rigor and to occur during the rigor process. The wrapping however was used to homogenise the meat and reduce variation. The wrapping, applied pre-rigor, stretches the muscle fibres so they have less opportunity to contract before they cool and the muscle hardens. A detailed description on the effect of wrapping was given by Devine *et al.* (1999).

## 5.4 Bloom

For the bloom time series colour images, the lean region was identified using the colour image processing method described above for each image in time, and the mean  $L$ ,  $a^*$  and  $b^*$  taken. In the corresponding hyperspectral images the lean region was identified as per the ECVA model.

Bloom over the first hour after slicing is due to the oxygenation of myoglobin, causing a change from purple to red (Hedrick *et al.*, 1994). At first none of the myoglobin is oxygenated so the rate of change is greatest. As the oxygenation proceeds less unoxxygenated myoglobin is available so the rate of change decreases. Assuming the rate of change proportional to the amount of unoxxygenated myoglobin then the oxygenation process follows an exponential decay process. A parametric exponential curve was fit to the  $a^*$  data over time,  $t$ , with the model

$$a^*(t) = a_0^* + \Delta a^*(1 - e^{-k(t-t_0)}), \quad (5.1)$$

where  $a_0^*$  is the initial value,  $\Delta a^*$  is the change in value during the blooming period,  $k$  is the rate of blooming and  $t_0$  is the slice time. A similar curve was obtained for  $b^*$ . To fit the blooming model to the colour image data a custom implementation of iterative least squares was used.

The  $L$  data, being luminance, is strongly influenced by surface scattering effects, hence is erratic over time so modelling the  $L$  values is not sensible. The  $L a^* b^*$  colour space does not completely separate the chroma from the intensity so  $a^*$  and  $b^*$  may be influenced by the erratic behaviour of  $L$ . PLS models for each of the six parameters  $a_0^*$ ,  $a_f^* = a_0^* + \Delta a^*$ ,  $k_a$ ,  $b_0^*$ ,  $b_f^* = b_0^* + \Delta b^*$  and  $k_b$  were evaluated using per animal cross validation.

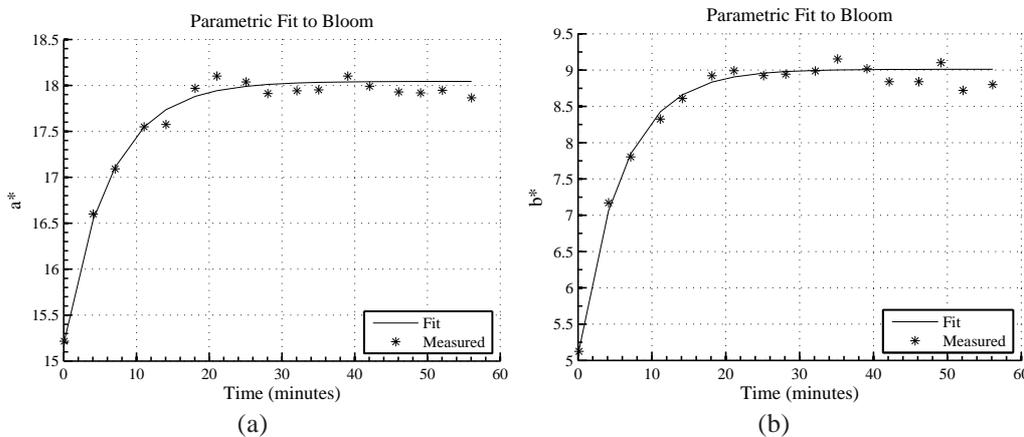


Figure 5.13: Example bloom curves from animal 6, left side, and the parametric curve fit to them. a) The  $a^*$  component and b) the  $b^*$  component over the blooming period are shown.

Unlike the slice images, the bloom colour images from the first day were salvageable. The bloom image acquisition was repeated on one slice over time so there was less variability to account for when processing the bloom images than the slice images. Furthermore many of the bloom images were taken when the lean had reached, or was near, the bright red colour characteristic of bloomed meat so segmentation of the lean region in the  $a^*$  plane was frequently easier in the bloom images than the slice images. Segmentation of the lean region and computation of the colour values were successful, which resulted in seven groups with two bloom curves per group.

The  $a^*$  and  $b^*$  bloom curves increased in a manner similar to an exponential approach to a maximum over the bloom period. Figure 5.13 shows the example bloom for animal 6, left side and the parametric curves fit. Over the fourteen muscles the average RMS error of the fit for  $a^* \pm \sigma$  is  $0.019 \pm 0.009$  and for  $b^* \pm \sigma$ ,  $0.03 \pm 0.01$ . The change in  $L$  over the bloom period was erratic and not consistent between samples. Some samples increased in  $L$ , others decreased, and some exhibited increase followed by decrease or vice versa. The most likely explanation is that the luminance was heavily influenced by specular reflection due to fluid on the meat face. This fluid was prone to change over the data collection period due to secretion, in which even drips formed and ran down the face of the meat. Such aggregation and movement of moisture affected the colour images much more than the hyperspectral imaging because the colour imager and the illumination were both near normal to the meat face whereas the spectrometer entrance slit was at about  $60^\circ$  to the surface normal of the meat, although influence on the hyperspectral images cannot be ruled out.

The lean spectra were identified by application of the lean pixel classifier (Section 5.2) and the mean lean spectrum taken for each spectral image. The spectrum from the initial time point, the final time point, the average over the bloom period and the difference between the initial and final spectra were compared to the computed  $a_0^*$ ,  $a_f^*$ ,  $k_a$ ,  $b_0^*$ ,  $b_f^*$  and  $k_b$  using PLS via per animal cross validation. The PLS models were built against each response individually. Crossvalidation was performed per animal, with the two samples per animal. Due to the small number of observations we use at most three PLS factors, but those models that reach three factors are considered tentative.

Table 5.5: Optimal number of factors, correlation and probability of spurious correlation for crossvalidation of the spectra against the bloom curve parameters. The numbers are formatted as: # Factors,  $R^2$  (p value).

	$a_0^*$	$a_f^*$	$k_a$	$b_0^*$	$b_f^*$	$k_b$
Initial spectra	2, 0.81(0.00)	2, 0.77(0.00)	×	3, 0.46(0.01)	2, 0.69(0.00)	×
Final Spectra	2, 0.81(0.00)	2, 0.83(0.00)	×	2, 0.46(0.01)	2, 0.74(0.00)	×
Difference Spectra	×	×	3, 0.40(0.02)	×	×	3, 0.31(0.04)

Table 5.5 lists the number of factors and correlations for the different spectra and responses after crossvalidation. Models marked with an  $\times$  indicate that no clear relation-

ship was observed, that is, there was no obvious optimal number of factors and a p test showed that for any number of factors the correlation was insignificant. The initial and final spectra resulted in a reasonable prediction of the initial and final  $a^*$  and  $b^*$  values and no prediction of the rate values. The prediction of the initial  $b^*$  values required three factors thus is tentative. The difference spectra resulted in weak correlation and poor tentative prediction of the bloom rate values. The most useful result here is that spectra taken immediately after cutting correlate with the final colour values. A larger scale study is required to investigate the prediction of final bloom on muscle prerigor and at various pertinent time points postrigor, in particular to study if spectra acquired at certain time points soon after slaughter can predict the final bloom value at time points consistent with the time of retail presentation.

Qiao *et al.* (2007b) and Andrés *et al.* (2008) both obtained good correlation with luminance,  $L$ , but not for the chroma,  $a^*$  and  $b^*$ . Our result is in contrast. Presumably in those studies measures were taken to prevent specularly contamination of the  $L a^* b^*$  measurements. Here, despite the blotting processed used to remove excess surface moisture, specularly made modelling  $L$  impractical. Obtaining significant correlations with  $a^*$  and  $b^*$  is encouraging and future studies with many more samples will allow us to expound on the relationship in more detail.

## 5.5 Tenderometry and pH Correlation

The correlation of the spectra per-wavelength with the pH and tenderometry data was calculated. The band that best described the response variables were chosen.

### 5.5.1 pH

Detrimental effects, such as an aging pH probe requiring more careful recalibration of the pH meter, impacted on the pH data from the first animal. None of the issues were insurmountable, but the first day's data was considered unreliable. Careful pH data acquisition left us with the remaining six animals data. Figure 5.14 shows the pH per slice, with vertical dashed lines to delineate each muscle. The data are ordered day two, left, slices 1–20; day two, right, slices 1–20; day three, left, and so on. Some of the muscles have drift in the pH measurements over the slices due to the issues surrounding the pH meter. The drift in measured pH is possibly increased by change in the meat pH during the data acquisition period (approximately two hours per muscle), however pH in meat normally decreases with time as aging occurs and the fact that some pH readings increase suggests issues with the meter is the more reasonable explanation. In the small number of animals represented there is a trend for the left side muscle to have lower pH than the right,

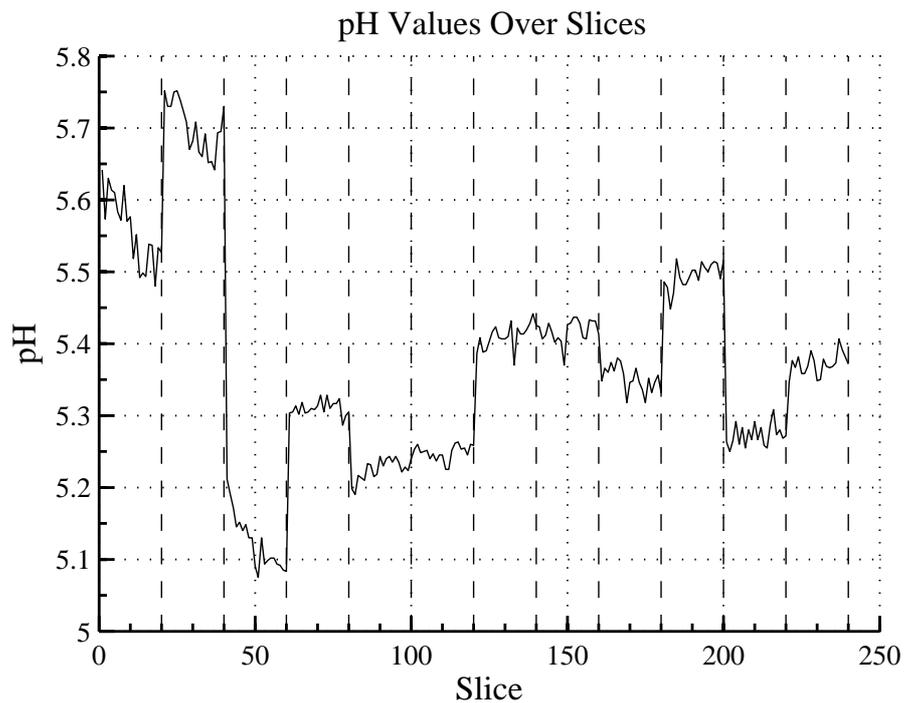


Figure 5.14: The pH readings down twelve LD muscles (the first animal is excluded). Vertical dashed lines separate one muscle from the next. The data are ordered per animal, left side followed by the right side. There is a tendency for the right side pH readings to be slightly higher than the left.

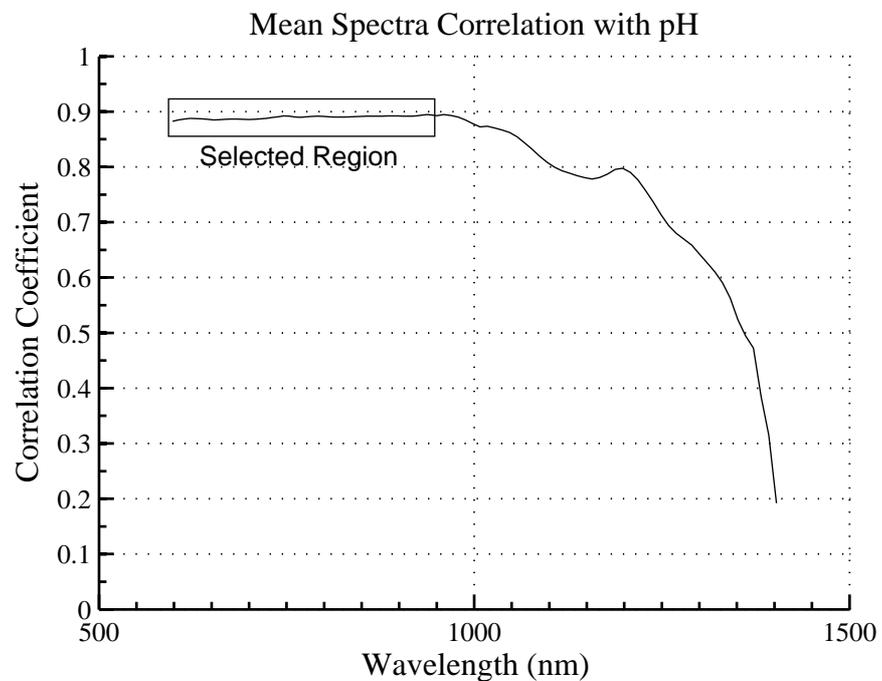


Figure 5.15: Correlation ( $r$  value) of each wavelength with the pH. The average of the selected region indicated is compared with the pH reference values.

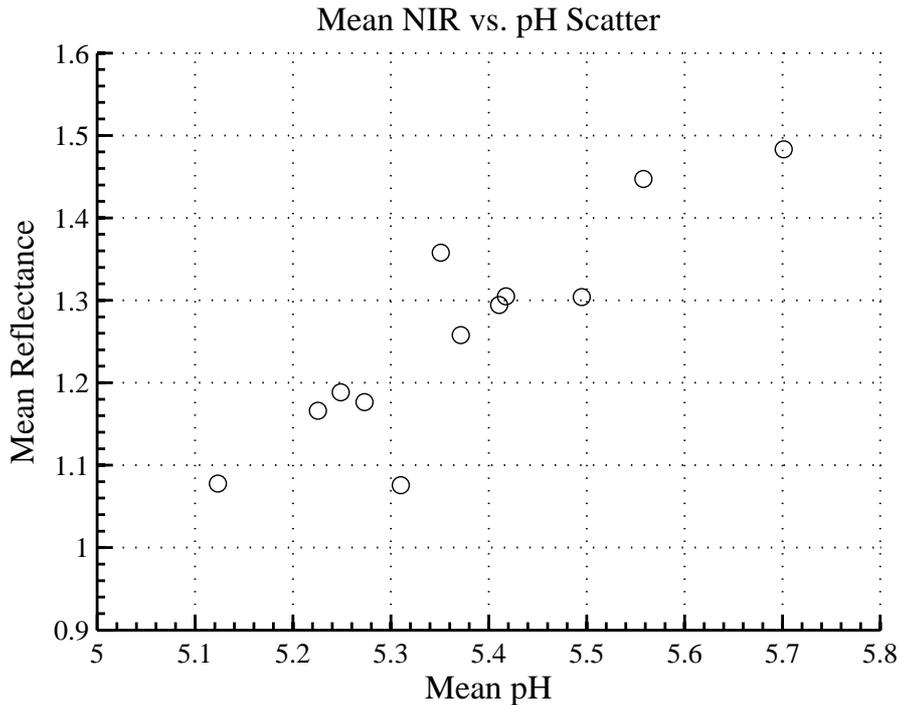


Figure 5.16: Scatter of the mean NIR reflectance in the band 598–949 nm against the pH.

possibly due to the carcass being hung by one leg at the Achilles tendon after slaughter causing more strain down one side than the other. More measurements on new carcasses are needed to confirm the trend of one side having lower pH than the other.

The Intra-LD spread in pH is small compared to inter-LD spread in our results. Effectively there are twelve unique pH measurements which is too small a set for any calibration-validation process. The lean spectra were identified using the lean pixel classifier described above. The mean lean spectrum and mean pH for each muscle was computed. Correlations were obtained per wavelength to observe the presence of a relationship between the spectra and the pH readings, and are shown in Figure 5.15. Strong correlation was observed between 600 and 950 nm, indicated as the selected region. The average of the spectra within 600 and 950 nm was taken and scatter plotted against the pH readings, as shown in Figure 5.16. A linear relationship is visible and has a correlation with  $R^2 = 0.797$  ( $p = 0.0001$ ). Fitting a line to the scatter in Figure 5.16 yields an RMS of the residuals of 0.068, about 10 % of the spread in the pH values.

The pH values here are in a very narrow range (5.08–5.75). Andersen *et al.* (1999) used 46 pork samples measured the day after slaughter, obtaining a pH range of 5.46–6.97 in the *semimembranosus* muscle which was subsequently used in calibration. Using PLS with leave one out crossvalidation Andersen *et al.* (1999) obtained a correlation of  $R^2 = 0.53$  on spectra of LD muscles over the band 1000–2630 nm. They also examined correlations per wavelength over 360–777 nm and 1000–2630 nm but did not obtain

the high and almost steady correlation over any of the band 598–949 nm that we did. This may be because of the different animal type examined (pig, not cow) or the time of measurement post mortem. Rosenvold *et al.* (2009) studied pH values over an extensive period of time post-mortem (from 1h to about 90h) with various treatments involving temperature while cooling and wrapping to influence the meat as rigor developed. As a result, they obtained a broad range of pH values 5.15–7.17. PLS was used to model the pH based on the spectra. Over the 253 samples used in the validation set they obtained  $R^2 = 0.84$ . The work of Rosenvold *et al.* (2009) is a much more comprehensive study than what is presented here, regardless the fact that a good correlation could be found in our data is in accordance with their result.

### 5.5.2 Tenderometry

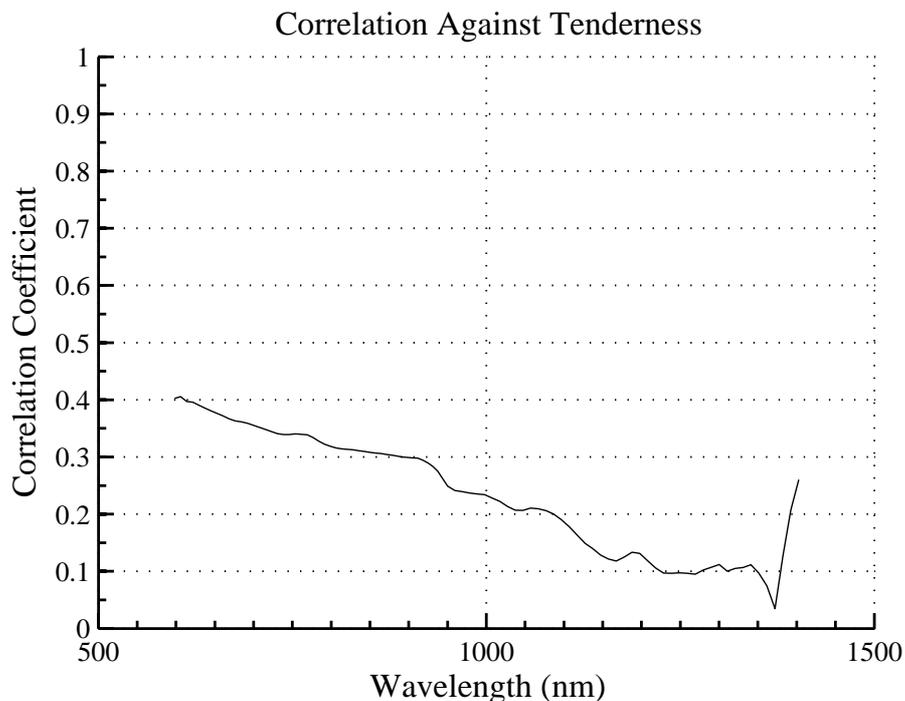


Figure 5.17: Correlation (r value) of tenderness readings against the wavelengths.

Figure 5.17 shows the correlation of the tenderness over the crossvalidation data against wavelength. There are no spectral regions that correlate well with the tenderometry readings. Any relationship between spectral effects and the tenderometry readings, if it exists at all, is not simple and is unlikely to be seen clearly or reliably with the small number of samples obtained. Five tenderometry measurements were taken of each steak, where the correlations in Figure 5.17 are of the wavelengths against the mean tenderometry for each muscle. The upshot is that each muscle's tenderometry reading has

an associated standard deviation, with the mean of the standard deviations to be 0.95 kgF. The spread in the average tenderometry readings is 3.49 kgF, so the ratio of the average standard deviation over the spread is 0.272. The uncertainty is nearly 30% of the spread. Clearly with the tenderometry measurement equipment available much greater spread in the tenderometry values is required before any useful correlation can be observed.

In the literature a mix of tenderness classification and calibration is reported. The purpose of investigating tenderness via spectroscopy is to find a way to grade the meat without damaging it. Grading for tenderness means assigning a label of tender or not tender, so classification is reasonable. We do not have enough samples to attempt classification, nor did we intend to, and the samples that we have do not fall into obvious groups. Rather we sought to measure quality features of the meat to inform the variography and pixel classification. Hildrum *et al.* (1994) and Naes and Hildrum (1997) obtained good classification and calibration of tenderness. They concluded that classification was better because calibration tended to return intermediate values, causing the more tender and tough samples to be misclassified.

Rødbotten *et al.* (2000), like us, did not obtain good prediction of tenderness ( $R^2$  values up to 0.46) but they were seeking a prediction of tenderness in the future, that is, they ask whether spectra taken early post-mortem can predict tenderness much later. Bowling *et al.* (2009) reported low correlation ( $R^2 \leq 0.23$ ) and observed that others who obtained good correlation had a broad range in tenderness values. Rødbotten *et al.* (2001) incorporated information regarding the treatment of the meat postmortem in performing prediction of tenderness in the future. They obtained better results in future prediction when treatment information was used than when it was not used ( $R^2$  in the range 0.27–0.69 without treatment information, 0.50–0.72 with treatment information). Rosenvold *et al.* (2009) obtained moderate prediction of tenderness ( $R^2 = 0.58$ ) but had a broad range of tenderness values (19–265 *N*) obtained from meat subjected to a range of treatments and measured at various times post-mortem. We have only used one treatment, that resulted in the narrow range of tenderness measurements stated above. Differing treatments of the meat potentially could have provided a broader range of tenderness values and produced better results. Varying treatments may also have yielded different results in the variography, potentially even drowning out the water effect observed. Regardless variography, along with tenderness, under different treatments is an interesting topic for future investigation.

# Chapter 6

## Conclusion and Future Work

This thesis presented the successful acquisition of hyperspectral images using source modulated spatial multiplexing. A projector system was developed that used a digital micromirror array (DMA) to project light patterns onto a sample. A spectrometer measured the spectral information reflected from the sample. The light patterns encoded the spatial information in such a way that the signal quality was improved over measuring each point on the sample at a time.

The theory for multiplexing was developed. In particular, discussion of all types of realistic error under multiplexing with the Hadamard matrices was presented. This discussion incorporated theoretical development that had gone before, encompassing the results into a single theoretical development. The effect of compressed sensing on multiplexing precision and some results concerning the use of strongly regular graph matrices were also discussed and developed.

The hyperspectral imager was tested and characterised for signal to noise ratio (SNR) behaviour. Aspects of the developed theory were verified, in particular, the reduction of the benefit in SNR due to Poisson noise was observed. The Hadamard matrices were also seen to be useful despite the SNR boost being reduced in certain high Poisson noise situations, especially when the main noise source (Poisson or additive Gaussian) is indeterminate or varying.

An important issue in all the work presented is that the hyperspectral imager is proof of concept in its design and implementation. As such funds were limited which impacted significantly in the selection of parts and equipment used. While the imager took good spectral images, it took a long period of time to do so (about 2.5 minutes to acquire a  $16 \times 16$  pixel image, up to about 40 minutes to acquire a  $64 \times 64$  pixel image) and the equipment is fragile. Better (and more expensive) parts are expected to reduce the acquisition time significantly and to enable development to improve the ruggedness of the device.

The imager was used to acquire spectra of beef *m. Longissimus Dorsi* from the right

and left sides of seven animals. The spectra were analysed for the classification power between lean and nonlean groups, variation down the length of the muscles, pH, bloom (change in colour after cutting) and tenderness (force required to sheer the meat after cooking). The proof of concept nature of the work meant that a large scale study was not sensible at this time. The positive results concerning the usability of the spectra produced by the system and the interesting features observed on the meat validate the approach. Furthermore, from a technical standpoint, further development of the imaging system and larger scale investigation of the meat is warranted.

## 6.1 Theory

The theory for multiplexing by the Hadamard matrices under additive and multiplicative systematic and random error was developed. Results from prior art were integrated into a single notational system under which the effect of the Hadamard H,G and S matrices on the different noise types were compared. For random additive noise the well known result surrounding noise reduction was derived. That is the H-matrix provides the best noise reduction, followed by the G-matrix and the S-matrix has the least noise reduction. For Poisson (random multiplicative) noise we confirmed that the S-matrix increased the noise variance, that the H-matrix has no effect on the average variance and it was shown that the G-matrix has the same noise increase as the S-matrix.

Systematic additive error (offset) was divided into two type, offset that is multiplexed and offset that is not multiplexed. It was observed that elimination of offset that is multiplexed is preferable, that is, any offset should be independent of the encoding. Additive offset that is independent of the encoding is corrected for before decoding and any random noise effects associated with the measurement used to make the correction is reduced in the subsequent decoding operation. If the offset is multiplexed, then the correction must be made after decoding and any random noise associated with the correction is not reduced.

The effect of systematic multiplicative drift at decoding was seen to be difficult to assess in the general sense. Because the error was mixed per entry into the multiplexing matrix the signal measured and the drift itself were required to compute the resultant error. Correction of multiplicative drift removed the bias but increased the overall noise level. Computation of the overall noise level was performed assuming additive and multiplicative noise were present in the signal and reference measurements. The reference measurements and the signal measurements were assumed independent, but the case of non independent measurements was recognised as possible and stated as requiring ad-hoc treatment. The correction of multiplicative drift produced new noise variance terms in addition to the terms that arise due to random noise. One of these terms showed a distinct

multiplexing disadvantage of the order of  $N$ , the size of the multiplexing matrix, however analysis in a later chapter with consideration of instrumental effects showed that this disadvantageous factor was mitigated by ensuring that attenuation through the reference beam is small. The H-matrix was better overall than the other two Hadamard matrices in terms of noise variance. Some variance terms in the S-matrix were smaller than for the G-matrix and others were larger, so neither had a distinct advantage.

The link between strongly regular graphs (SRGs) and Hadamard H and G-matrices was explored. The link between SRGs and the S-matrix was established by Ratner *et al.* (2007) herein the connection to the other Hadamard matrices is explored. The SRG matrices have useful properties for multiplexing when Poisson noise is present, especially when the relative Poisson and additive noise levels are known before multiplexing takes place. One construction procedure for the S-matrix is to start from the H-matrix, from which the G-matrix and subsequently the S-matrix are derived. Starting with the SRG matrix, this construction procedure was reversed to produce H-like and G-like matrices. The properties of the new matrices were examined via the matrix trace and eigenvalue structure. It was found that the G-like matrix has properties closer to the Hadamard H-matrix. The H-like matrix, while providing a good multiplex boost, was not quite as good as the G-like matrix. We take this as evidence that while the S-matrix shares many properties with a SRG matrix under certain conditions, they are not exactly equivalent. Regardless, SRG matrices form attractive multiplexing patterns for certain situations.

Compressed sensing (CS) theoretical precision using a random encoding matrix was examined against the number of measurements. It was shown that the random error in a CS reconstruction depends only on the inversion of the acquisition pattern acting on the random error in the acquisition. For a large random matrix, as the number of rows increases the spread in the eigenvalues increases. From the eigenvalue structure it is inferred that the precision decreases as the number of measurements increases. A numerical simulation was performed in which a sparse signal was CS ‘multiplexed’ using a random acquisition pattern to which random error was added. The signal was subsequently reconstructed using basis pursuit for a range of the number of measurements and the SNR computed for each reconstruction. The theoretical SNR was computed assuming that random noise at acquisition was the only error source. The SNR in CS reconstruction was small for very few measurements in accordance with others analyses. The SNR increased until it met the theoretical SNR from our model and then decreased closely tracking to the theoretical prediction.

The Wilcoxon and the area under the receiver operator curve (ROC AUC) are two equivalent statistics for assessing the power of a classification system to assign data to groups. The effect of error in the reference in computing the Wilcoxon was derived and discussed. A direct link between the two statistics was established by beginning with the

the definition of the AUC and deriving a result which was seen to match the definition of the Wilcoxon. The resultant equation was then modified and analysed under the presence of error in the reference which manifested as the transference of members between groups. The Wilcoxon was found to be reduced by error in the reference. Given an estimate of the number of data transferred, we showed that an estimate of the Wilcoxon without error could be computed. Numerical simulations of computing the Wilcoxon on data with varying numbers of data transferred between groups verified the theoretical prediction of the effect of error in the reference on the Wilcoxon.

## 6.2 Hardware

A hyperspectral imager was built and tested. The imager utilised source modulated multiplexing using a custom digital image projector with a broadband light source. The projector imaged multiplexing patterns onto the sample and a single point spectrometer integrated light from the sample. The light source projected ‘on-off’ patterns of squares onto the sample with each square corresponding to one pixel. Changing the patterns over time while taking spectra built a linear transformation of the spectral image pixels. Later decoding returned the individual spectral pixel responses. The spatial domain was multiplexed while the spectral domain was examined by the spectrometer, so each spectral band, or wavelength, was multiplexed in parallel. The error types considered in the theory chapter were interpreted for the optical system. Additive random error occurred in the spectrometer and multiplicative random noise occurred as photon fluctuations from the light source that followed Poisson statistics. Systematic additive error occurred as background illumination and the baseline offset in the spectrometer while systematic multiplicative error occurred as light source drift. The theory was reworked to incorporate hardware specific considerations such as the important inclusion of the optical attenuation effect.

A complement encoding scheme was used to remove additive offset. Assuming that the encoding matrix (Hadamard or otherwise) had both positive and negative entries, then the pixels at locations that corresponded to the positive entries were illuminated, then immediately following were illuminated according to the negative entries, and spectra were taken for each illumination. The positive and negative encodings are complementary, hence we call the process complement encoding. Since the positive and negative encodings were taken in quick succession there is minimal chance for drift in the background offset to occur, so the additive offsets in both spectra were the same and subtraction of the negative encoding from the positive encoding performed the necessary correction. Since the encodings were complementary no desired spectral response from the sample was lost. For multiplexing matrices that have no negative entries, such as the Hadamard

S-matrix and pointwise acquisition, a background measurement was required for each encoding. The background measurement was acquired by setting the encoding to uniformly zero (no illumination of the sample) and taking a spectrum. To perform the correction the background spectrum was subsequently subtracted from the encoding spectrum.

A reference beam monitored the light source intensity over the acquisition period. The reference beam measurements were subsequently used for light drift correction. A reference spectrum was taken with and subsequently divided from every encoding spectrum. Due to physical size constraints the acquired raw reference beam measurements actually were the combination of the reference beam and sample so to obtain the reference beam the sample spectrum without the reference beam was subtracted from the raw reference. The subtraction operation caused the sample and reference measurements to be dependent. The theory surrounding the effect on random error of reference beam correction was reworked to account for the lack of independence which increased the noise variance slightly. There was a disadvantageous term in Hadamard multiplexing that occurred due to correction of multiplicative drift. Accounting for optical effects showed that if the reference beam optics attenuate much less than the sample optics then the combined optical effect mitigates the disadvantage. Also the rise in noise variance due to the lack of independence between the sample and reference measurements was greatly mitigated by the relatively low attenuation by the reference beam optics.

In decoding the image spectra the optical attenuation effect remained after the corrections and were applied. A reference white tile was used to obtain an image of the attenuation effect. The reference white tile was assumed to have a very high and flat reflectance profile over the spectral band of interest. Dividing the sample image by the white tile image removed the optics effect. The only error remaining after this final correction was the random noise. The theoretical noise variance was computed for the white tile corrected spectra. It was found that a low light level through the sample optics resulted in an increase in noise variance, but was mitigated by a sufficiently intense light source. The overall noise variance was approximately doubled by the white tile correction. The theoretical analysis lead to the commonsense conclusion that more light from the light source and less attenuation by the optics ultimately lead to less noise in the final corrected spectra.

Hadamard matrix and compressed sensing patterns were implemented. SRG matrix multiplexing was not implemented because the relative additive and multiplicative noise level varied over the wavelengths, so a multiplexing pattern designed to give the optimal boost where the photon noise was greatest would not have been optimal where the light level, hence the signal, was low. Instead, prompted by the theory, the Hadamard multiplexing was expected to provide a good multiplexing advantage at wavelengths of low light level where it was most needed and that the advantage would be smaller where the

light level, hence the SNR, was already good. Since the H-matrix does not reduce Poisson noise variance on average, the varying SNR boost was interpreted as a preferential noise reduction at lower light levels which is demonstrably useful when the main noise source is indeterminate or varying, which is the case with our system.

Hadamard G-matrix and H-matrix encodings were implemented. These Hadamard encoding matrices used have both positive and negative entries so complement encoding was used. Complement encoding, with its inherent advantages and disadvantages, is one of the key features of the work herein. S-matrix encoding does not require complement encoding and provides no theoretical advantage over the other Hadamard matrices so was not used here. The complement encoding is used to remove background illumination and the baseline offset in the spectrometer. To do the same for S-matrix encoding requires measurement of the baseline with no added value in terms of acquiring information about the sample. Complement encoding performs the background correction using measurements that have spectral information from the sample, a qualitative reason why the G-matrix encoding is better than the S-matrix with regards to signal independent additive noise.

In a multiplexing situation where the correction of systematic additive error is not required then complement encoding is not necessary, and the subtraction operation that doubles the random noise may even render complement encoding disadvantageous. One must assess the benefits of the encoding system for the particular situation. We believe that Hadamard H-matrix encoding using complement encoding is the best approach for the hardware configuration described herein. There was a background offset inherent in the spectrometer; zero illumination did not return a zero spectrum and the baseline point could drift over time. Background illumination due to light sources not related to the spectrometer system and stray light from the spectrometer light source both contributed to the additive offset. Also, as stated above, the H-matrix, which can only be implemented on our system by complement encoding, provides the useful preferential boost.

Hyperspectral images of simple objects were taken and the system was assessed by examination of the signal to noise ratio of the spectra. Visual examination of the images of the simple objects showed that the images produced are sensible. Processing of the spectra using SNV, SVD and thresholding segmented the images successfully, demonstrating that in simple objects the spectral imager produces usable spectra.

SNR experiments were conducted to examine the SNR boost over pointwise imaging for the G and H-matrices. For the G-matrix multiplexing a  $N = 255$  matrix was used to capture  $17 \times 15$  pixel images. The SNR boost reached the maximum of  $\sqrt{255/2} = 11.29$  at the highest and lowest wavelengths, which is where the light output from the bulb and the detector sensitivity were both lowest. The SNR boost dropped greatly where the light bulb output, hence the photon noise, was greatest and the detector sensitivity, hence

relative instrument noise, was least. The SNR boost was least where the SNR is greatest, and was greatest where the SNR was least, indicating that the SNR boost was delivered where it was most needed.

For the H-matrix encoding a  $N = 256$  matrix was used to capture  $16 \times 16$  pixel images. Like with the G-matrix multiplexing, the maximum SNR boost of  $\sqrt{256} = 16$  was reached at the lowest and highest wavelengths and the minimum SNR boost occurred where the SNR was greatest. The H-matrix maximum SNR boost, which occurs when additive noise dominates the noise processes, was approximately  $\sqrt{2}$  times that of the G-matrix boost as predicted. The general shape of the SNR boost was the same for the two encoding matrices but with the important difference that the photon noise did not reduce the boost relative to the maximum possible boost to the same degree that the G-matrix encoding did, as predicted. For a complement encoding scheme H-matrix multiplexing is superior to the G-matrix.

Compressed sensing (CS) was implemented using a  $N = 256$  matrix of randomly assigned ones and negative ones to perform  $16 \times 16$  pixel spectral imaging. With CS it was necessary to perform SNR experiments to assess measurement precision and error experiments to assess accuracy. Both accuracy and precision varied with the number of measurements taken. As the number of measurements increased the SNR decreased as predicted and the accuracy increased as indicated by the literature, thus there is a tradeoff between accuracy and precision. The software used to perform the CS reconstruction has a noise aware reconstruction option. The noise aware reconstruction performed very poorly indicating that the real data acquired did not satisfy the assumptions in the algorithm in accounting for random fluctuations.

## 6.3 Application

To demonstrate the applicability of the imaging system a larger and more difficult problem than imaging simple objects was attempted. Spectral volumetric mapping of beef *m. longissimus dorsi* (LD) muscles was performed. LDs are long and after collection they were rolled producing a long tubular shape. Volumetric mapping was performed by slicing the meat and taking hyperspectral images of each slice. Concurrent with the hyperspectral imaging colour images were acquired and subsequently processed to identify the visible content of each spectral pixel. The pH of each slice was measured and tenderometry readings were taken on a steak cut from each muscle.

Classification of the spectra as lean or nonlean was trialled. Air spectra were easily identified and removed as part of the data cleaning process, thus the air spectra were not processed as part of the lean pixel classification. The main visible components left were lean and fat. Some connective tissue was visible in the colour images but the image

processing did not segment out the connective tissue regions well. No preprocessing, such as SNV or MSC, was used on the spectra as such methods remove gross intensity which was considered an important spectral feature. Extended canonical variates analysis (ECVA) was used to discriminate between the lean and nonlean groups. ECVA is the application of PLS to solve a Fisher linear discriminant analysis problem when there is a large degree of covariance between the wavelengths in spectral data. Being based on PLS, ECVA produces factors that can be examined individually. A three factor model was chosen from the testing stage for use in validation. The first factor did not produce strong separation of the lean and nonlean groups for all the muscles used in the training data set, indicating that gross intensity difference was not always the strongest discriminating factor. The other two factors produced good separation between the groups over the different muscles. The AUC of a ROC curve, equivalently the Wilcoxon statistic, was used to assess the classification power. The AUC of validation was 0.9515. In other words the ECVA model returned a higher score for the lean than the nonlean in about 95% of the validation spectral pixels indicating very good classification power.

The technique of variography was used to examine the nature of the variation in the spectra down the length of the muscles. Variography examines the similarity on average (correlation) between data at different points for varying distances between points. The shape of the variogram is indicative of the type of variation in the samples from which the data (spectra) are taken. The variography indicated that the variation within a given muscle is small, but increased with distance down the length. There was a periodic effect in the variography that corresponded with the spatial period of the vertebrae that the LDs were cut from. There are two possible explanations for the periodicity in the variograms. The first is that a spatially periodic removal of connective tissue corresponding with the vertebrae that is produced when the LD is cut from the carcass. Connective tissue retains water, so where the connective tissue is removed from the lean more seepage is likely. The second explanation is a possible strain effect on the muscle fibres when the carcass is hung postslaughter. Each vertebrae could pull on the muscle causing a periodic effect.

A bloom curve was obtained for each muscle using both colour and hyperspectral imaging. The colour images were processed to obtain CIE  $L^*a^*b^*$  values for each time point. Exponential decay parametric curves were fit to the  $a^*$  and  $b^*$  values to obtain the initial values, final values and the rate of change of the bloom. The  $L$  values were erratic and unusable, most likely due to specularities in the colour images. The lean spectra in the bloom spectral images were identified using the ECVA classifier and averaged to obtain one spectrum per muscle per time point. Using SIMPLS and per animal cross validation the initial spectra, final spectra and difference spectra were related to the parameters of the bloom curve. The initial and final spectra both related well to both the initial and final bloom parameters. The difference spectra related moderately well to the bloom rate

parameter, but the model should be considered tentative. A much larger scale study is required to assess if initial spectra can predict the bloom rate and if spectra taken at some time point soon after slaughter can predict the final colour at a time point consistent with retail presentation.

Variation of pH within each muscle was small compared to inter-muscle variation, so the average pH for each muscle was taken, giving twelve data points (a very small number). The wavelength band from 598 to 949 nm was found to correlate well with the pH and the average absorbance in this band returned  $R^2 = 0.7973$  with  $p = 0.0001$ . Given the small number of measurements correlation with such a low p-value is a good result, yet many more measurements are needed to establish a calibration of pH from spectral reflectance. Other limiting factors are the small spread in pH values and the use of one time point post slaughter. The small spread in pH values is undoubtedly related to the use of one time point, but is further exacerbated by the collection and handling protocol which was intended to optimise the tenderness in the meat.

Prediction of tenderometry failed in this study. The principle cause of difficulty was the small range of tenderometry values caused by the handling procedure. No useful correlations were observed. Varying the handling procedure and timepoints for measurement was identified as necessary to broaden the range of tenderometry readings. The literature that we compared our work to contained a mix of classification and calibration approaches to the prediction of tenderness via NIR. The work presented here is not intended to create a data set suitable for classification.

## 6.4 Future Directions

The foremost concern of future direction is the improvement of the light projector system. The heart of the imager is the DMA chip in the projector. The DMA used is from an off-the-shelf digital data projector and is limited in speed by the associated electronic hardware. Better suited hardware to drive the DMA will improve data acquisition speed significantly. The optical system is proof of concept in nature and is fragile. Investment into the physical design will result in a more robust optical system. The visible domain and the high wavelength end of the NIR region were too noisy. Research into improving the spectral breadth of the light source output will improve the width of the spectral region not dominated by noise. The most obvious improvement to make is to use a tungsten halogen bulb with a higher colour temperature than what was used herein (2800 Kelvin). A higher colour temperature has greater emission in the visible domain. The overall intensity output increases with colour temperature, so care will be required to ensure that the DMA chip is not harmed if a similar optical scheme to what is described here is employed.

The actual physical construction shown in Chapter 4 is fragile and nonportable. A future construction should be more robust and self contained. Ultimately desired is a single box unit containing the projector optics and spectrometer, with a view window on top to place samples on for imaging. In a new construction both the reference beam *and* the light to the sample should be chopped so that the covariance between the sample and reference measurements is zero. Purchasing a DLP kit from Texas Instruments that includes the chip and driver hardware and software is critical for the successful implementation of a more robust unit. The modified data projector approach taken here to get access to a working DMA, while sufficient at proof of concept level, is the main cause of fragility in the system. Like the hardware presented in this thesis, the next iteration in development would involve pulling together the appropriate parts from ‘off the shelf’ sources. Further iterations may require the development of specialised electronic driver circuitry for optimised speed and synchronisation of the modulated illumination and data acquisition process.

Calibration transfer, that is the use of a model built from data from one spectrometer on data from other spectrometers, is always an issue. We anticipated that if a two dimensional sensor like a CCD was used then there would even be a transfer problem within the data generated by the one spectrometer, thus a single point spectrometer with one sensor per wavelength was used. Regardless if development of this system progresses then transfer of calibrations between different units built will no doubt be necessary.

We have identified a tradeoff between precision and accuracy in compressed sensing. Some outstanding issues remain. The precise interpretation of the tradeoff between accuracy and precision and the relationship with overfitting verses wrong model error that can occur in least squares regression problems are issues for further consideration. The theoretical effect on random noise, computable from the eigenvalues of the random acquisition system, assumes that the acquisition matrix has entries drawn at random from a Gaussian distribution. The numerical example given meets the assumption of entries drawn at random from a Gaussian, however the acquisition of real data was performed using an acquisition system of randomly assigned ones and negative ones. The theoretical connection of the effect on precision between the two similar but different types of acquisition matrices need to be established. The results were as expected but with one exception. The denoising option in the basis pursuit technique did not improve the reconstruction, in fact the opposite occurred. We can only conclude that the data did not meet the assumptions of the basis pursuit denoising technique, however the reason why is not currently known and should be investigated further.

The theoretical result concerning the Wilcoxon statistic when there is error in the reference has room for further development. In particular we have assumed that the reference method and the new method under examination are independent, hence are uncorrelated.

We havenot studied in detail the effect of correlation between the reference method and a new classification technique. Independence between reference and the new method means that data misclassified by the reference method are transferred between groups without bias from the perspective of the new method. Correlation between the two methods means that there is bias from the perspective of the new method. For example, if the two methods are positively correlated then the misclassification is more likely to occur at the boundary between the two classes in the new method. Conversely, negatively correlated methods mean that the misclassification is more likely to occur at the opposite extreme values away from the boundary in the new method. Precisely how to mathematically model the effect of error in the reference method on the new method when there is correlation is not known at this time. Modelling independent of any particular distribution is preferred, but it may be that assumptions about the type of distribution are required to facilitate analysis. There is also scope for numerical modelling with non-Gaussian distributions to broaden the verification of the theory.

The application of the hyperspectral imager is limited in scope. Repetition of the meat spectra analysis with many more samples is necessary to establish commercially viable calibrations and clearer understanding of the variability studies. The variography will benefit greatly from the inclusion of many more samples. Spectra for variography need to be taken of meat that has undergone varying handling protocols and at different time points post slaughter. The bloom analysis requires many more samples to establish calibrations that are not considered tentative in nature. The work presented in this thesis can inform the design of experiment analysis to establish how many samples are necessary. The work of Rosenfold *et al.* (2009) answers many of the questions surrounding pH prediction that might arise from this work. Regardless pH measurement is always useful and informative in the analysis of meat, so subsequent comparison of pH readings with spectra is sensible. The cause of the difficulties in predicting tenderometry reading, namely the lack of breadth of measurements that we obtained, were identified. Varying handling protocols and time points for variography will also cause a greater range of tenderometry readings. Causing variation in the tenderness readings will also allow the classification of spectra into different tenderness groups to be investigated.



## References

- Andersen, J., C. Borggaard, A. Rasmussen, and L. Houmøller. Optical measurements of pH in meat. *Meat Science*, **53**, pp. 135–141 (1999).
- Andrés, S., A. Silva, A. Soares-Pereira, C. Martins, A. Bruno-Soares, and I. Murray. The use of visible and near infrared reflectance spectroscopy to predict beef *M. longissimus thoracis et lumborum* quality attributes. *Meat Science*, **78**, pp. 217–224 (2008).
- Baraniuk, R., M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, **28(3)**, pp. 253–263 (2008).
- Barker, M. and W. Rayens. Partial least squares for discrimination. *Journal of Chemometrics*, **17(3)**, pp. 166–173 (2003).
- Barnes, R., M. Dhanoa, and S. Lister. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, **43**, pp. 772–777 (1989).
- Barrett, H., C. K. Abbey, and E. Clarkson. Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions. *JOSA A*, **15(6)**, pp. 1520–1535 (1998).
- Becker, E. Fourier transform spectroscopy. *Science*, **178(4059)**, pp. 361–368 (1972).
- Bowling, M., D. Vote, K. Belk, J. Scanga, J. Tatum, and G. Smith. Using reflectance spectroscopy to predict beef tenderness. *Meat Science*, **82**, pp. 1–5 (2009).
- Brainard, D. H. The psychophysics toolbox. *Spatial Vision*, **10**, pp. 443–436 (1997).
- Candès, E. The restricted isometry property and its implications for compressed sensing. *Theory of Signals/Mathematical Analysis*, **346(9-10)**, pp. 589–592 (2008).
- Candes, E., J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, **52(2)**, pp. 489–509 (2006).

- Candès, E. and T. Tao. Decoding by linear programming. *IEEE Transactions on Information Theory*, **51(12)**, pp. 4203–4215 (2005).
- Chen, S. S., D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, **20(1)**, pp. 33–61 (1998).
- Cortes, C. and M. Mohri. AUC optimization vs. error rate minimization. *Advances in Neural Information Processing Systems*, **16** (2004). MIT Press.
- Damaschini, R. Limitation of the multiplex gain in Hadamard transform spectroscopy. *Pure and Applied Optics: Journal of the European Optical Society Part A*, **2(3)**, pp. 173–177 (1993).
- Davis, D. S. Multiplexed imaging by means of optically generated kronecker products: 1. the basic concept. *Applied Optics*, **34(7)**, pp. 1170–1176 (1995).
- DeVerse, R., R. Hammaker, and W. Fateley. Hadamard transform raman imagery with a digital micro-mirror array. *Vibrational Spectroscopy*, **19(2)**, pp. 199–186 (1999).
- DeVerse, R., R. Hammaker, and W. Fateley. Realization of the Hadamard multiplex advantage using a programmable optical mask in a dispersive flat-field near-infrared spectrometer. *Applied Spectroscopy*, **54(12)**, pp. 1751–1758 (2000).
- Devine, C., N. Wahlgren, and E. Tornberg. Effect of rigor temperature on muscle shortening and tenderisation of restrained and unrestrained beef *m. longissimus thoracis et lumborum*. *Meat Science*, **51**, pp. 61–72 (1999).
- Dhanoa, M., S. Lister, R. Sandersona, and R. Barnes. The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. *Journal of Near Infrared Spectrosc*, **2**, pp. 43–47 (1994).
- Donoho, D. Compressed sensing. *IEEE Transactions on Information Theory*, **52(4)**, pp. 1289 – 1306 (2006).
- Fateley, W., R. Hammaker, R. DeVerse, R. Coifman, and F. Geshwind. The other spectroscopy: demonstration of a new de-dispersion imaging spectrograph. *Vibrational Spectroscopy*, **29**, pp. 163–170 (2002).
- Fateley, W. G., R. M. Hammaker, and R. A. DeVerse. Modulations used to transmit information in spectrometry and imaging. *Journal of Molecular Structure*, **550-551**, pp. 117–122 (2000).

- Fernández, F., J. Vadillo, F. E. J. Kimmel, R. Zare, N. Rodriguez, M. Wetterhall, and K. Markides. Effect of sequence length, sequence frequency, and data acquisition rate on the performance of a Hadamard transform time-of-flight mass spectrometer. *Journal of the American Society for Mass Spectrometry*, **12(12)**, pp. 1302–1311 (2001).
- Gehm, M. E., R. J. Brady, R. M. Willett, and T. J. Schulz. Single-shot compressing spectral imaging with a dual-disperser architecture. *Optics Express*, **15(21)**, pp. 14013–14027 (2007).
- Gill, P., W. Murray, D. Ponceleón, and M. Saunders. Solving reduced KKT systems in barrier methods for linear and quadratic programming. Technical Report SOL 91-7, Stanford University, Stanford, CA, USA (1991).
- Golay, M. Multi-slit spectrometry. *Journal of the Optical Society of America*, **39**, pp. 437–444 (1949).
- González-Martín, I., C. González-Pérez, J. Hernández-Méndez, and N. Alvarez-García. Determination of fatty acids in the subcutaneous fat of Iberian breed swine by near infrared spectroscopy (NIRS) with a fibre-optic probe. *Meat Science*, **65**, pp. 713–719 (2003).
- Gy, P. Sampling of discrete materials II. quantitative approach—sampling of zero-dimensional objects. *Chemometrics and Intelligent Laboratory Systems*, **74**, pp. 25–38 (2004a).
- Gy, P. Sampling of discrete materials III. quantitative approach—sampling of one-dimensional objects. *Chemometrics and Intelligent Laboratory Systems*, **74**, p. 39–47 (2004b).
- Hanley, J. and B. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**, pp. 29–36 (1982).
- Hanley, Q. and T. Jovin. Highly multiplexed optically sectioned spectroscopic imaging in a programmable array microscope. *Applied Spectroscopy*, **55(9)**, pp. 1115–1123 (2001).
- Hanley, Q., P. Verveer, D. Arndt-Jovin, and T. Jovin. Three-dimensional spectral imaging by Hadamard transform spectroscopy in a programmable array microscope. *Journal of Microscopy*, **197**, p. 5 (2000).
- Hanley, Q., P. Verveer, and T. Jovin. Optical sectioning fluorescence spectroscopy in a programmable array microscope. *Applied Spectroscopy*, **52(6)**, pp. 783–789 (1998).

- Hansen, S., T. Hansen, M. Aaslyng, and D. Byrne. Sensory and instrumental analysis of longitudinal and transverse textural variation in pork *longissimus dorsi*. *Meat Science*, **68**(4), pp. 611–629 (2004).
- Harwit, M. and N. J. Sloane. *Hadamard Transform Optics*. Academic Press, New York, USA (1979).
- Hassler, K., T. Anhut, and T. Lasser. Time-resolved Hadamard fluorescence imaging. *Applied Optics*, **44**(35), pp. 7564–7572 (2005).
- Haupt, J., R. Castro, and R. Nowak. Compressed sensing in noisy imaging environments. *Proceedings of SPIE 6065*, p. 606507 (2006).
- Haupt, J. and R. Nowak. Signal reconstruction from noisy random projections. *IEEE Transactions on Information Theory*, **52**(9), pp. 4036–4048 (2006).
- Hedrick, H. B. ., E. D. Aberle, J. Forrest, M. D. Judge, and R. Merkel. *Principles of Meat Science*, chapter 6, pp. 126–129. Kendall/Hunt, Iowa, USA, third edition (1994).
- Henderson, H. V. and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, **23**(1), pp. 53–60 (1981).
- Hildrum, K., B. Nilsen, M. Mielnik, and T. Naes. Prediction of sensory characteristics of beef by near-infrared spectroscopy. *Meat Science*, **38**, pp. 67–80 (1994).
- Hornbeck, L. Deformable mirror spatial light modulators. *SPIE Critical Reviews Series*, **1150**, pp. 86–102 (1989).
- Huang, H., H. Yu, H. Xu, and Y. Ying. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of Food Engineering*, **87**, pp. 303–313 (2008).
- Janz, J., J. Aalhus, M. Dugan, and M. Price. A mapping method for the description of Warner–Bratzler shear force gradients in beef *Longissimus thoracis et lumborum* and *Semitendinosus*. *Meat Science*, **72**(1), pp. 79–90 (2006).
- de Jong, S. Simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, **18**(3), pp. 251–263 (1993).
- MacFarlane, P. and J. Marer. An apparatus for determining the tenderness of meat. *Food Technology*, **6**, pp. 134–135 (1966).

- Maggioni, M., G. L. Davis, F. J. Warner, F. B. Geshwind, A. C. Coppi, R. A. DeVerse, and R. R. Coifman. Hyperspectral microscopic analysis of normal, benign and carcinoma microarray tissue sections. In: *SPIE Optical Biopsy VI*, volume 6091, pp. 92–100 (2006).
- Martens, H., S. A. Jensen, and P. Geladi. Multivariate linearity transformation for near infrared reflectance spectra of meat. In: *Application Spectroscopy, Proceedings of the Nordic Symposium, Applied Statistics*, pp. 235–267. Stockholm Forlag Publication, Stavanger, Norway (1983).
- Martens, H., J. Pram Nielsen, and S. Balling Engelsen. Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry*, **75(3)**, pp. 394–404 (2003).
- Miller, K. S. On the inverse of the sum of matrices. *Mathematics Magazine*, **54(2)**, pp. 67–72 (1981).
- Naes, T. and K. Hildrum. Comparison of multivariate calibration and discriminant analysis in evaluating nir spectroscopy for determination of meat tenderness. *Applied Spectroscopy*, **51(3)**, pp. 350–357 (1997).
- Nguyen, L. and B. A. J. Young. All-optical CDMA with bipolar codes. *Electronic Letters*, **31(6)**, pp. 469–470 (1995).
- Nitzsche, G. and R. Riesenberger. Noise, fluctuation, and Hadamard-transform spectrometry. *Fluctuations and Noise in Photonics and Quantum Optics*, **5111(1)**, pp. 273–282 (2003).
- Nørgaard, L., R. Bro, F. Westad, and S. B. Engelsen. A modification of canonical variates analysis to handle highly collinear multivariate data. *Journal of Chemometrics*, **20(8-10)**, pp. 425–435 (2006).
- Osborne, B., T. Fearn, and P. Hindle. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*. Longman Scientific and Technical, Harlow, UK, second edition (1993).
- Otsu, N. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, **9(1)**, pp. 62–66 (1979).
- Ott, L. and W. Mendenhall. *Understanding Statistics*, chapter 14. Duxbury Press, Boston, Massachusetts, USA (1985).

- Pearson, K. Mathematical contributions to the theory of evolution – on a form of spurious correlation which may arise when indices are used in the measurement of organs. In: *Proceedings of The Royal Society*, volume 60, pp. 489–498 (1897).
- Pelli, D. G. The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, pp. 437–442 (1997).
- Petersen, L. and K. H. Esbensen. Representative process sampling for reliable data analysis – a tutorial. *Journal of Chemometrics*, **19(11-12)**, pp. 625–647 (2005).
- Qiao, J., M. Ngadi, N. Wang, C. Gariépy, and S. Prasher. Pork quality and marbling level assessment using a hyperspectral imaging system. *Journal of Food Engineering*, **83**, pp. 10–16 (2007a).
- Qiao, J., N. Wang, M. Ngadi, A. Gunenc, M. Monroy, C. Gariépy, and S. Prasher. Prediction of drip-loss, pH, and color for pork using a hyperspectral imaging technique. *Meat Science*, **76**, pp. 1–8 (2007b).
- Ratner, N., Y. Y. Schechner, and F. Goldberg. Optimal multiplexed sensing: bounds, conditions and a graph theory link. *Optics Express*, **15(25)**, pp. 17072–17092 (2007).
- Rødbotten, R., B.-H. Mevikand, and K. Hildrum. Prediction and classification of tenderness in beef from non-invasive diode array detected nir spectra. *Journal of Near Infrared Spectroscopy*, **9**, pp. 119–210 (2001).
- Rødbotten, R., B. Nilsen, and K. Hildrum. Prediction of beef quality attributes from early post mortem near infrared reflectance spectra. *Food Chemistry*, **69**, pp. 427–436 (2000).
- Rosenvold, K., E. Micklander, P. W. Hansen, R. Burling-Claridge, M. Challies, C. Devine, and M. North. Temporal, biochemical and structural factors that influence beef quality measurement using near infrared spectroscopy. *Meat Science*, **82**, pp. 379–388 (2009).
- Sierra, V., N. Aldai, P. Castro, K. Osoro, A. Coto-Montes, and M. Oliván. Prediction of the fatty acid composition of beef by near infrared transmittance spectroscopy. *Meat Science*, **78**, pp. 248–225 (2008).
- Silverstein, J. W. The smallest eigenvalue of a large dimensional Wishart matrix. *The Annals of Probability*, **13(4)**, pp. 1364–1368 (1985).
- Sloane, N., T. Fine, P. Phillips, and M. Harwit. Codes for multiplex spectrometry. *Applied Optics*, **8(12)**, pp. 2103–2106 (1969).

- 
- Smola, A. J. and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, **14(3)**, pp. 199–222 (2004).
- Spudich, T., C. Utz, J. Kuntz, R. DeVerse, R. Hammaker, and D. McCurdy. Potential for using a digital micromirror device as a signal multiplexer in visible spectroscopy. *Applied Spectroscopy*, **57(7)**, pp. 733–736 (2003).
- Streeter, L., G. Burling-Claridge, and M. Cree. Image processing of meat images for visible/near infrared spectroscopy reference. In: *Image and Vision Computing New Zealand 2006*, pp. 275–280. Gt Barrier Island, New Zealand (2006a).
- Streeter, L., G. Burling-Claridge, M. Cree, and R. Künnemeyer. Visible/near infrared hyperspectral imaging via spatial illumination source modulation. *Journal of Near Infrared Spectroscopy*, **15(6)**, pp. 395–399 (2007).
- Streeter, L., G. Burling-Claridge, M. Cree, and R. Künnemeyer. Comparison of Hadamard imaging and compressed sensing for low resolution hyperspectral imaging. In: *Image and Vision Computing New Zealand 2008*, pp. 1–6. Christchurch, New Zealand (2008a).
- Streeter, L., G. Burling-Claridge, M. Cree, and R. Künnemeyer. Reference beam method for source modulated Hadamard multiplexing. *Proc SPIE*, **6812**, pp. 68160J1–10 (2008b).
- Streeter, L., G. Burling-Claridge, M. Cree, and R. Künnemeyer. Optical full Hadamard matrix multiplexing and noise effects. *Applied Optics*, **48(11)**, pp. 2078–2085 (2009).
- Streeter, L., R. Burling-Claridge, and M. J. Cree. Colour image processing and texture analysis on images of porterhouse steak meat. In: *Image and Vision Computing New Zealand 2005*, pp. 398–343. Dunedin, New Zealand (2005).
- Streeter, L., P. Hansen, R. Burling-Claridge, M. Cree, R. Künnemeyer, and C. Devine. Estimation of visible fat content in beef m. longissimus using near infrared spectroscopy. In: *52nd International Congress of Meat Science and Technology*, p. 671. Dublin, Ireland (2006b).
- Sylvester, J. Thoughts on inverse orthogonal matrices, simultaneous sign successions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers. *Philosophical Magazine*, **34**, pp. 461–475 (1867).

- Takhar, D., J. Laska, M. Wakin, M. Duarte, D. B. S. Sarvotham, K. Kelly, and R. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. *Proc. SPIE 6065*, p. 606509 (2006).
- Tøgersen, G., J. Arnesen, B. Nilsen, and K. Hildrum. On-line prediction of chemical composition of semi-frozen ground beef by non-invasive nir spectroscopy. *Meat Science*, **63**, pp. 515–523 (2003).
- Tsaig, Y. and D. Donoho. Extensions of compressed sensing. *Signal Processing*, **86(3)**, pp. 533–548 (2006).
- Verveer, P., Q. Hanley, P. Verbeek, L. V. Vliet, and T. Jovin. Theory of confocal fluorescence imaging in the programmable array microscope (PAM). *Journal of Microscopy*, **189(3)**, pp. 192–198 (1998).
- Vincent, L. Morphological grayscale reconstruction in image analysis: Applications and efficient algorithms. *IEEE transactions on image processing*, **2(2)**, pp. 176–201 (1993).
- Wagadarikar, A., R. John, R. Willett, and D. Brady. Single disperser design for compressive, single-snapshot spectral imaging. *Proc. SPIE*, **6714** (2007).
- Wakin, M., J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk. An architecture for compressive imaging. In: *2006 IEEE International Conference on Image Processing*, pp. 1273–1276. Atlanta, Georgia, USA (2006a).
- Wakin, M., J. Laska, M. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. Kelly, and R. Baraniuk. Compressive imaging for video representation and coding. In: *Picture Coding Symposium*. Beijing, China (2006b).
- Westbrook, C. *MRI at a glance*. Blackwell Science, Oxford, UK (2002).
- Wilhelmi, G. and F. Gompf. Binary sequences and error analysis for pseudo-statistical neutron modulators with different duty cycles. *Nuclear Instruments and Methods*, **81(1)**, pp. 36–44 (1970).
- Wold, H. Path models with latent variables: the NIPALS approach. In: *H.M. Blalock et al. Quantitative sociology: international perspectives on mathematical and statistical modeling*, pp. 307–375. Academic Press, New York, USA (1975).
- Wuttig, A. Optimal transformations for optical multiplex measurements in the presence of photon noise. *Applied Optics*, **44(14)**, pp. 2710–2719 (2005).
- Xia, J., E. Berg, J. Lee, and G. Yao. Characterizing beef muscles with optical scattering and absorption coefficients in vis-nir region. *Meat Science*, **75**, pp. 78–83 (2007).

Xia, J., A. Weaver, D. Gerrard, and G. Yao. Heating induced optical property changes in beef muscle. *Journal of Food Engineering*, **84**, pp. 75–81 (2008).