

Some pages of this thesis may have been removed for copyright restrictions.

If you have discovered material in Aston Research Explorer which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our [Takedown policy](#) and contact the service immediately (openaccess@aston.ac.uk)

**Advances in Characterisation,
Calibration and Data Processing
Speed of Optical Coherence
Tomography Systems**

Janarthanan Rasakanthan

Doctor of Philosophy

ASTON UNIVERSITY

February 2014

© Janarthanan Rasakanthan, 2014

Janarthanan Rasakanthan asserts his moral right to be identified as the author of this thesis.

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgment.

Abstract

Advances in Characterisation, Calibration and Data Processing Speed of Optical Coherence Tomography Systems

Janarthanan Rasakanthan

Doctor of Philosophy

February 2014

Aston university

This thesis describes advances in the characterisation, calibration and data processing of optical coherence tomography (OCT) systems.

Femtosecond (fs) laser inscription was used for producing OCT-phantoms. Transparent materials are generally inert to infra-red radiations, but with fs lasers material modification occurs via non-linear processes when the highly focused light source interacts with the materials. This modification is confined to the focal volume and is highly reproducible.

In order to select the best inscription parameters, combination of different inscription parameters were tested, using three fs laser systems, with different operating properties, on a variety of materials. This facilitated the understanding of the key characteristics of the produced structures with the aim of producing viable OCT-phantoms. Finally, OCT-phantoms were successfully designed and fabricated in fused silica. The use of these phantoms to characterise many properties (resolution, distortion, sensitivity decay, scan linearity) of an OCT system was demonstrated.

Quantitative methods were developed to support the characterisation of an OCT system collecting images from phantoms and also to improve the quality of the OCT images. Characterisation methods include the measurement of the spatially variant resolution (point spread function (PSF) and modulation transfer function (MTF)), sensitivity and distortion.

Processing of OCT data is a computer intensive process. Standard central

processing unit (CPU) based processing might take several minutes to a few hours to process acquired data, thus data processing is a significant bottleneck. An alternative choice is to use expensive hardware-based processing such as field programmable gate arrays (FPGAs). However, recently graphics processing unit (GPU) based data processing methods have been developed to minimize this data processing and rendering time. These processing techniques include standard-processing methods which includes a set of algorithms to process the raw data (interference) obtained by the detector and generate A-scans. The work presented here describes accelerated data processing and post processing techniques for OCT systems.

The GPU based processing developed, during the PhD, was later implemented into a custom built Fourier domain optical coherence tomography (FD-OCT) system. This system currently processes and renders data in real time. Processing throughput of this system is currently limited by the camera capture rate. OCT-phantoms have been heavily used for the qualitative characterization and adjustment/fine tuning of the operating conditions of OCT system. Currently, investigations are under way to characterize OCT systems using our phantoms.

The work presented in this thesis demonstrate several novel techniques of fabricating OCT-phantoms and accelerating OCT data processing using GPUs. In the process of developing phantoms and quantitative methods, a thorough understanding and practical knowledge of OCT and fs laser processing systems was developed. This understanding leads to several novel pieces of research that are not only relevant to OCT but have broader importance. For example, extensive understanding of the properties of fs inscribed structures will be useful in other photonic application such as making of phase mask, wave guides and microfluidic channels. Acceleration of data processing with GPUs is also useful in other fields.

Key words: Optical coherence tomography (OCT), Graphics Processing Unit (GPU) data processing, femtosecond laser (fs) inscription, characterisation and calibration phantoms

Dedicated to my parents.

Acknowledgements

First and for most I would like to take this opportunity to thank my supervisor Dr. Kate Sugden for giving me a chance to work on this project, and for all the support she has given to me.

I would also like to thank Dr. Peter. H. Tomlins and Dr. Peter. D. Woolliams for sharing their knowledge of OCT and for the valuable opportunity they have given to me to work with them. I have learned a great deal about practical aspects of OCT, measurement and characterization methods from them.

I am also grateful for the financial support by National Physical Laboratory (NPL) via an Engineering and Physical Sciences Research Council (EPSRC) Case studentship award.

Thanks to Dr. Kaiming Zhou and Dr. Mykhaylo Dubov, I have acquired a great deal of knowledge on the femtosecond inscription. I would also like to thank Graham Lee and Graham Smith, with whom I have worked to produce the OCT-phantoms.

I am also grateful to the advice given by Dr. Vladimir Mezentsev, Dr. John A.R. Williams and Prof. David Webb.

My sincere thanks also go to Mathew Tedaldi at NPL for giving me access to OCT in his lab, whenever I needed and Andy Main for all the help he has given me at the start of my PhD.

I am grateful to Hetel Patel, Mandana Braghen for working with me.

I am grateful to Jiangling Li (Alice) and Lutful Khan for all the coffee and tea. Time in Aston would not have been enjoyable without these people. It has been

a great pleasure to have their company on this journey.

Many thanks should also go to Helen Yard, Sandra Smith, Yuen Chu, Bert Biggs and Andy Abbot for their help and support along the way with anything from drilling holes to purchasing and travel.

Finally, I would like to thank my parents and sisters for their love and support during my PhD.

Contents

1	Introduction	24
1.1	Background and Motivation	24
1.2	Thesis Structure	27
2	Background	31
2.1	Optical Coherence Tomography	31
2.1.1	Introduction	31
2.1.2	Theory	35
2.2	Femtosecond Inscription	40
2.2.1	Introduction	40
2.2.2	Theory	41
2.3	Parallel Processing	50
2.3.1	Introduction	50
2.3.2	CUDA	53
2.4	Conclusion	61
3	OCT phantom requirements	63
3.1	Resolution	64
3.1.1	Rayleigh Criterion	64
3.1.2	Lateral Resolution	65
3.1.3	Axial Resolution	65
3.1.4	Modulation transfer function	67

CONTENTS

3.2	Method	68
3.3	Point spread function	70
3.4	Focusing conditions	72
3.5	Sensitivity	74
3.6	Distortion phantom	76
3.7	Oversampling and Smoothing	78
3.7.1	One dimension	78
3.7.2	Two dimensional data	79
3.8	Deconvolution	80
3.8.1	Theory	81
3.8.2	Tiled deconvolution	84
3.8.3	Quantification of improvement by deconvolution	85
3.9	Conclusion	86
4	Experimental set-up	88
4.1	Femtosecond laser	88
4.1.1	Lower repetition rate system: 800 nm	88
4.1.2	Lower repetition rate system: 1026 nm	90
4.1.3	High repetition rate system: 800 nm	93
4.1.4	Sample preparation and alignment	93
4.1.5	Writing direction	95
4.1.6	Stage motion	95
4.2	Optical Coherence Tomography	97
4.2.1	Michelson diagnostic OCT (MD-OCT)	97
4.2.2	Santec OCT	99
4.2.3	Obtaining Data	99
4.2.4	Slicing angle	100
4.2.5	Point selection	102
4.2.6	Selecting a sub-region	105
4.3	Conclusion	106

CONTENTS

5	Parameter Space Test	108
5.1	<i>Low-Rep 1026 nm</i> system	109
5.1.1	Preliminary study	109
5.1.2	Extended parameter space study	119
5.1.3	Material differences	125
5.2	Lower Repetition <i>800 nm</i> system	130
5.2.1	Fused silica glass	132
5.2.2	Polymethyl methacrylate (PMMA)	134
5.2.3	Discussion	139
5.3	High repetition <i>800 nm</i> system	140
5.4	Discussion	141
5.5	Conclusion	142
6	Development of characterisation phantoms	145
6.1	PMMA - Single layer	145
6.2	Fused Silica - Multi layer	148
6.2.1	General design	148
6.2.2	PSF phantom	150
6.2.3	Sensitivity phantom	152
6.2.4	Distortion phantom	155
6.2.5	Lateral resolution phantom	157
6.2.6	Discussion	160
6.3	Conclusions for both PMMA and Fused silica	161
7	OCT system characterisation using OCT phantoms	162
7.1	Resolution	162
7.1.1	Method	162
7.1.2	Results and Discussion	163
7.1.3	Summary	166
7.2	Lateral resolution	167
7.3	Sensitivity	170

CONTENTS

7.4	Sensitivity drop	173
7.5	Distortion	174
7.6	Conclusion	174
8	Deconvolution	175
8.1	Method	175
8.2	Results	177
8.2.1	Blurred image	177
8.2.2	Mean PSF	178
8.2.3	Varying (Local) PSF	179
8.3	Discussion	182
8.4	Conclusion	185
9	Data Processing	186
9.1	Interpolations	187
9.1.1	Linear interpolation	188
9.1.2	Lagrangian Interpolation	188
9.2	Method	191
9.3	GPU implementation	193
9.3.1	The kernels configuration	196
9.4	Methods	199
9.4.1	Processing rate	199
9.5	Results and Discussion	201
9.6	Conclusion	208
10	Real Time Optical Coherence Tomography	210
10.1	Introduction	210
10.1.1	Description of the OCT system	211
10.1.2	Description of the personal computer (PC)/software model	215
10.2	Conclusion	216

CONTENTS

11 Conclusion	219
11.1 Conclusion	219
11.2 Future work	223
11.2.1 Parameter space study	224
11.2.2 Phantom design and fabrication	224
11.2.3 Quantitative methods for characterization	225
11.2.4 Software/Algorithm developments	225
A Gaussian fit	246
B Reconstructing the refractive index	250

List of Figures

2.1	Time domain optical coherence tomography (TD-OCT) set-up. . .	32
2.2	Spectral domain optical coherence tomography (SD-OCT) set-ups: spectrometer based optical coherence tomography (SB-OCT) (top) and swept-source optical coherence tomography (SS-OCT) (bottom). 34	
3.1	Airy disk separation: resolved (left), Rayleigh criterion (middle) and unresolved (right).	65
3.2	Simulated axial resolution (coherence length) as function of varying full width at half maximum (FWHM) spectral width for three different central wavelengths (800, 1300 and 1500 nm).	66
3.3	The simulated results of sinusoidal (MTF) and top-hat (contrast transfer function (CTF)) response functions.	67
3.4	Simulation results for detected particle size, obtained with equa- tion (3.3.3), by OCT with various resolutions, for particles with vary- ing size.	71
3.5	Longitudinal and transverse size of the beam as function of NA for central wavelengths 800 nm and 1026 nm.	74
3.6	Example of calculating the sharpness.	86
4.1	Experimental set-up of LR800 fs laser system.	90
4.2	Experimental set-up of the <i>Low-Rep 1026 nm</i> system.	91
4.3	Measured <i>Low-Rep 1026 nm</i> power at the objective as function of relative polariser position.	92

LIST OF FIGURES

4.4	Example of sample alignment in <i>Low-Rep 800 nm fs</i> system. . . .	94
4.5	Example of writing direction (bottom-up) used for inscription of the OCT-phantoms.	95
4.6	Point spacing as function of translation speed for three repetition rate (top). Microscope images of example structures inscribed by <i>Low-Rep 800 nm</i> (bottom).	96
4.7	The galvanometer scan (red line) used to obtain B-scans and linear stage used to obtain volumetric data.	98
4.8	Example cross section when structures are scanned at an angle other than normal (i.e. wedge).	100
4.9	The radius of the cross section (semi-major axis) as a function of scan angle α	101
4.10	Step 1 of the point identification procedure: Extracting region of interest around the identified points.	103
4.11	Example of a 2D Gaussian blurring kernel.	103
4.12	Step 2 of points identification procedure: enhancing image.	104
4.13	Step 3 of points identification procedure: identifying the centroid of each points in the selected region.	104
4.14	Regions selected for analysis.	105
4.15	Example of a sub-section extracted around a structure.	106
5.1	A schematic of the femtosecond inscribed OCT calibration phantom showing the paired lines of varying power increasing from left to right after initial location lines [1].	110
5.2	a) Microscope image of three pairs of lines showing a variation in the feature width with varying inscription power b) QPm measurements of two lines showing the contrast between above material threshold void creation (above) and below material threshold index change.	111

LIST OF FIGURES

5.3	Variation of measured apparent defect size with laser power setting used taken at a depth of $75 \mu m$ from the surface, as measured using an optical microscope [1]. ¹	112
5.4	OCT B-scan of the preliminary parameter space test (PST) sample, scaled in logarithmic intensity	113
5.5	Confocal microscope (CM) images of the cross section of modified regions after etching in HF acid for various times.	114
5.6	CM (top) and scanning electron microscope (SEM) (bottom) images of the cross section of the cleaved and chemically etched structures inscribed in fused silica sample.	115
5.7	Axial and lateral width of the structures measured using an OCT and a CM [1].	116
5.8	CM images (grey) of the cleaved, etched phantom with corresponding OCT cross sections (colour) shown on a intensity linear scale.	117
5.9	CM images of the cleaved and etched cross-section of the lines inscribed using 2.5% of the total energy.	118
5.10	A schematics of the extended parameter space sample. Each of the gray region in (a) represents lines inscribed at 20 different energies as shown in (d).	120
5.11	Microscope images of the structures showing variation in the feature width with varying energy, varying speed and varying depth.	121
5.12	Variation of apparent lateral size for the structures as varying laser energy, speed and depth.	122
5.13	OCT cross-section of the sample in logarithmic logarithmic scale.	123
5.14	The FWHM axial (left) and lateral (right) width of the damaged cross-sections.	124
5.15	Photograph (left) and schematic layout (right) of parameter space test inscribed in FeO_2 particle embedded resin phantom.	127
5.16	The lateral size of the structures inscribed with varying energy for constant speed and depth.	128

LIST OF FIGURES

5.17	The lateral size of the structures inscribed versus depth for constant speed and energy.	128
5.18	OCT results of the first 10 lines from a PST phantom fabricated by <i>Low-Rep 1026 nm</i>	129
5.19	The FWHM axial and lateral width of the first layer, obtained from OCT B-scan data.	130
5.20	Layout of the PST phantom for <i>Low-Rep 800 nm</i> system.	131
5.21	Microscope images of the structures inscribed by <i>Low-Rep 800 nm</i>	133
5.22	Lateral width, obtained from the optical microscope, for a) varying depth at constant energy and speed, b) varying energy at constant depth.	133
5.23	Microscope images of the structures inscribed by <i>Low-Rep 800 nm</i> system.	135
5.24	Microscope images of the structures inscribed in PMMA with high speed value of $v = 1 \text{ mm/s}$ and a high energy value of $E = 178 \text{ nJ}$ (a-f) and with a low energy value of $E = 35.6 \text{ nJ}$ (g-h).	136
5.25	B-scan of the PMMA PST sample, where structures were inscribed with varying energy, depth for constant scanning speed.	137
5.26	Microscope measurement of the structures inscribed by the <i>Low-Rep 800 nm</i> inside PMMA. Lateral size variation with a) varying inscription energy, b) varying depth and c) speed.	138
5.27	FWHM axial and lateral width of the structures, inscribed in PMMA, measured from OCT B-scans.	139
5.28	Microscope images of the wave-guides inscribed in doped-BK7 glass by a <i>High-Rep 800 nm</i> system.	140
5.29	B-scan OCT images show sample of the waveguides inscribed with <i>High-Rep 800 nm fs</i> system.	141
6.1	The single layer PMMA phantom.	147
6.2	Three dimensional rendering of PMMA single layer phantom	148

LIST OF FIGURES

6.3	Schematics of the overall layout (front view) and the photograph of the OCT calibration and characterisation phantoms.	149
6.4	PSF phantom	151
6.5	Three dimensional rendering of PSF phantom.	152
6.6	Sensitivity phantom.	154
6.7	Three dimensional rendering of sensitivity phantom.	155
6.8	Distortion phantom.	156
6.9	Three dimensional rendering of grid phantom.	157
6.10	Lateral resolution phantom.	159
6.11	Three dimensional rendering of the lateral resolution phantom. . .	160
7.1	Histograms for the FWHM of (a) lateral and (b) axial PSF.	163
7.2	Measured FWHM lateral (blue) and axial (green) PSF as a function of axial position.	164
7.3	The variation of the a) axial and b) lateral PSF as a function of axial lateral position in the B-scan.	165
7.4	B-scan data of the lateral resolution pattern obtained by a) <i>MD-OCT systemsystem</i> and b) Santec-OCT system.	168
7.5	<i>En-face</i> images (top) and oversampled line profile (bottom) extracted from set 1 (left) and 2 (right) of the lateral resolution pattern.	169
7.6	OCT images of structures inscribed with varying energy at $100 \mu m$ below the surface of fused silica used as sensitivity phantom.	171
7.7	Integrated intensity from sensitivity sample.	172
7.8	a) The B-scan of the PSF pattern and b) the measured intensity of the points across the B-scan plane.	173
7.9	Identified intersections (a) and measured distortion tensor of part of the select points (b) from the sample data set.	174
8.1	Example of calculating average PSF of a tile.	176

LIST OF FIGURES

8.2	B-scan of blurred image (top) and region used for qualitative comparison. Blue rectangle represent the region that can be deconvolved using the spatially PSF data obtained using PSF pattern, which is split into tiles indicated by red rectangles. The green rectangle represent the region (show in bottom) that used for qualitative comparison.	177
8.3	Histograms of (a) axial and (b) lateral resolution prior to deconvolution.	178
8.4	Mean PSF used for deconvolution of each sub section.	179
8.5	The results after applying deconvolution using Lucy-Richardson (LR) (top row) and Wiener filter (WF) (bottom row) routines using mean PSF. B-scan after deconvolution (a,e); mesh plots (b,f); axial FWHM (c,g); lateral FWHM (d,h).	180
8.6	Statistical distribution of local contrast (LC)(a,e), contrast-to-noise-ratio (CNR)(b,f), signal to noise ratio (SNR) (c,g) and sharpness K (d,h) for LR(top row) and WF (bottom) routines using mean(global) PSF.	181
8.7	Local (mean) PSF corresponding to each tile, that was used for deconvolving that tile. PSF within each tile is assumed to be invariant.182	
8.8	The results after applying deconvolution using LR (top row) and WF (bottom row) routines using tiled PSF (i.e. PSF is locally invariant within each tile). B-scan after deconvolution (a,e); mesh plots (b,f); axial FWHM (c,g); lateral FWHM (d,h).	183
8.9	Statistical distribution of LC(a,e), CNR(b,f), SNR(c,g) and sharpness K (d,h) for LR(top row) and WF (bottom) routines using local PSF	184
9.1	Illustrating the Lagrangian interpolation method.	190
9.2	Schematics of the GPU based data processing flowchart.	192
9.3	The flowchart for validating GPU results.	192

LIST OF FIGURES

9.4	Copy data from CPU memory to GPU Constant memory at the initialisation state.	194
9.5	Data flowchart during the pre-processing.	195
9.6	Flow chart of post-processing, where dot-dot lines represent the alternative paths.	196
9.7	Example kernel configuration showing a 3×5 blocks arranged in a two dimension (2D) grid, where each grid is composed of 8×8 threads	197
9.8	Flowchart for timing a) Compute Unified Device Architecture (CUDA) kernels and b) data transfer.	200
9.9	Tesla C1060 execution time line.	201
9.10	Maximum and standard deviation of the processing speed achieved with concurrent copy and transfer method for 100,000 A-scans. . .	202
9.12	The processing time of different kernels measured on Tesla 1060 and GT240 GPUs.	203
9.11	Processing rate for 140 different runs for Tesla C1060 and GeForce 9000M	204
9.13	Processing time taken for different tasks (kernel) as percentage of total processing time on different GPU and for different interpolations.	205
9.14	Processing speed with varying number of A-scans in batch, for Tesla GPU.	207
9.15	Processing rate achieved, using concurrent copy and execution, for varying number of A-scans. stream-0 is the first stream. Total is the total processing time.	208
10.1	Optical layout of the RT-OCT.	211
10.2	Mirror rotation and optical scan range	214
10.3	The morphological visualisation of mucocele obtained using RT-OCT. a) Top view and b) internal view.	217
10.4	Images of the PSF phantom obtained by RT-OCT.	217

LIST OF FIGURES

A.1	Simulated input data (a) and the peak estimated by fitting a 2D Gaussian (b). Input data was simulated with the fit parameters $\{1, 11, 11, 3, 3, 1\}$ and with a 20 dB noise.	249
B.1	Illustrating the propagation of light through a cylindrical structure of radius r	251
B.2	Shows a typical example serious of images taken in focus (middle), and $2\ \mu\text{m}$ either side of the focus (top and bottom).	253
B.3	A typical a) phase image and b) the reconstructed refractive index of the waveguides inscribed by <i>High-Rep 800 nm</i>	254

List of Tables

2.1 GPU hardware memory hierarchy. R/W: Read and write; R: read only.	55
4.1 Properties of the translational stage	91
4.2 Properties of the OCT systems	97
5.1 Summary of laser parameters used in the PST study in both low repetition femtosecond laser systems.	109
5.2 Line number in the phantom and their corresponding inscription energy.	120
5.3 Details of the parameter space phantom study using <i>Low-Rep 1026 nm</i> laser system.	126
5.4 Inscription parameters used for the PMMA and fused silica PST artefacts employing the <i>Low-Rep 800 nm</i> femtosecond system. The energy range chosen here was the the results of preliminary investigation conducted with a large range of values.	131
8.1 The summary of quantitative measure of the deconvolved tiled mean PSF	178
8.2 The summary of quantitative measure of the deconvolved tiled mean PSF	179
8.3 The summary of quantitative measure of the deconvolution Tiled Local PSF.	182

LIST OF TABLES

9.1	Details of the GPUs used in this study	194
9.2	Max, Min, mean and standard deviation of the processing rates for three GPUs	203
9.3	Average pre-processing rate on Tesla for varying number of threads per block.	206

Acronyms

1D one dimension.

2D two dimension.

3D three dimension.

ALU Arithmetic logical unit.

API application-programming interface.

AR anti reflection.

CCD charge-coupled device.

CM confocal microscope.

CMOS complementary metal oxide semiconductor.

CNR contrast-to-noise-ratio.

CPU central processing unit.

CTF contrast transfer function.

CUDA Compute Unified Device Architecture.

CUFFT CUDA fast Fourier transform.

CWD compute work distribution.

DAC digital-to-analogue converter.

DAQ data acquisition.

DDR3 SDRAM double data rate type **fs** femtosecond.

three synchronous dynamic random access memory.

DFT discrete Fourier transform.

DLL dynamic-link library.

DOF depth of field.

DPU double precision unit.

DRAM dynamic random access memory.

EPSRC Engineering and Physical Sciences Research Council.

FD Fourier domain.

FD-OCT Fourier domain optical coherence tomography.

FDML Fourier domain mode locked.

FFT fast Fourier transformation.

FFT fast Fourier transform.

FFTW Fastest Fourier Transform in the West (an FFT library).

FPGA field programmable gate array.

ACRONYMS

FT Fourier transform.	OM optical microscope.
FWHM full width at half maximum.	OPD optical path difference.
GPU graphics processing unit.	OpenCL OpenCL.
GUI graphical user interface.	OpenGL OpenGL.
I/O input/output.	OSM overlap-save method.
InGaAs Indium gallium arsenide.	OTF optical transfer function.
LC local contrast.	PC personal computer.
LCI low coherence interferometry.	PCIe Peripheral Component Interconnect Express.
LR Lucy-Richardson.	PMMA Polymethyl methacrylate.
LSF line spread function.	PSD power spectral density.
MD-OCT Michelson diagnostic OCT.	PSF point spread function.
MPA multi-photon absorption.	PSF-map point spread function map.
MPI multi-photon ionisation.	PST parameter space test.
MRI magnetic resonance imaging.	PTX parallel thread execution.
MTF modulation transfer function.	QPM quantitative phase microscopy.
NA numerical aperture.	ROI region of interest.
NI National Instruments.	SB-OCT spectrometer based optical coherence tomography.
NIR near infra-red.	SD spectral domain.
NLPI non-linear photo-ionisation.	SD-OCT spectral domain optical coherence tomography.
NPL National Physical Laboratory.	SDK software development kit.
NSR noise to signal ratio.	SEM scanning electron microscope.
NVCC NVIDIA CUDA compiler.	SFU special function unit.
OAM Overlap-Add method.	SIMD single instruction, multiple data.
OCDR optical coherence domain reflectometry.	SIMT single instruction, multiple threads.
OCT optical coherence tomography.	
OFDI optical frequency domain imaging.	

ACRONYMS

SLED super luminescent diode.

SM streaming multiprocessor.

SNR signal to noise ratio.

SPMD single programme, multiple data.

SS-OCT swept-source optical coherence tomography.

SVD singular value decomposition.

TD time domain.

TD-OCT time domain optical coherence tomography.

TPC texture/processor cluster.

USA United States of America.

WF Wiener filter.

WLI white light interferometry.

Chapter 1

Introduction

1.1 Background and Motivation

Optical coherence tomography (OCT) is a non-invasive three-dimensional imaging system that is capable of producing high-resolution *in-vivo* images. OCT is derived from optical coherence domain reflectometry (OCDR), which in turn is based on white light interferometry (WLI) [2, 3]. OCDR is a one-dimensional range finding technique that was developed to find faults in optical fibres and soon its potential for probing eyes and other biological tissues has been realised [4–6]. OCDR was extended by adding lateral beam scanning to provide cross sectional images of objects investigated. A cross sectional image of eye tissue was first demonstrated in 1991 by Huang et al., [7].

The first generation of OCT images used low coherence interferometry (LCI) with time domain detection, hence named time domain optical coherence tomography (TD-OCT). By continuously sweeping the reference arm in TD-OCT, echo delays in the backscattered light were detected. The next generation of OCT relied on Fourier Domain detection techniques, hence named Fourier domain optical coherence tomography (FD-OCT) systems. The FD-OCT systems exhibit speed and sensitivity advantages over TD-OCT. The dramatic improvements in acquisition speed by OCT made it possible, to acquire volumetric data in real time,

with a record dynamic range. Fercher et al., [8], demonstrated Fourier domain interferometry applied to measurement of eye length. State of the art modern systems applying similar technology can achieve an acquisition of few hundred thousand A-scans per second [9–13].

The drawback of FD-OCT is that it requires significantly more data processing to handle the large amount of information recorded and historically this has incurred a significant time overhead. Therefore, the use of graphics processing unit (GPU) for OCT data processing is of interest since they are optimized for performing a high number of parallel procedures and memory bandwidth. GPUs require parallel programming algorithms to exploit their performance power, for which specific programming platforms are available. Compute Unified Device Architecture (CUDA) is one of development suite created by NVIDIA (Santa Clara, California) that comes with tools, drivers and many libraries that are necessary to programme NVIDIAs GPUs for general-purpose computation.

OCT systems have been approved for use in clinical trials in Japan, United States of America (USA) and Europe. For OCT systems to be used effectively in clinical diagnosis, standardisation methods are required to assess their performance. As OCT systems become more widespread, verifying their performance against a reference standard becomes vital. A reference standard is required to quantify variations amongst different parameters as well as to assess the performance degradation of OCT system over time. Therefore reliable measurement phantoms and processing techniques are essential to quantify the OCT system performance, as well as to implement qualitative control procedures.

Regulatory bodies such as European Council require validation of the imaging systems performance using standard phantoms to ensure the uniformity of the reported performance between institutions and over time. Lack of standard calibration and characterization using known, well-calibrated phantoms leads to uncertainties in the reported results. Therefore, validating the performance of the OCT using a globally accepted standard would increase confidence in the reported performance of the systems. Such standards require easy to use, re-

producible phantoms. Ideally, multiple performance parameters should be characterized using a single phantom or set of data.

This standardization can be implemented using highly accurate and reproducible phantoms for calibration at both installation and throughout the lifetime of a system. Standard phantoms are necessary for validating and optimizing the performance of the system by the vendors and for routine quality control checks in the clinics. The main advantage of using standard phantoms is intra-laboratory consistency of the reported systems performance.

Important parameters of the OCT systems include resolution, sensitivity, contrast, distortions, linearity and imaging depth.

The three dimension (3D) point spread function (PSF) of the system can be estimated by measuring an infinitesimally small structure (i.e. structures of sizes much smaller than the resolution along each coordinate).

The sensitivity of an OCT system can be measured using micro-structures with varying refractive index.

The distortion introduced by an OCT can be measured by imaging phantoms made of parallel lines or grids.

Femtosecond (fs) lasers can write highly reproducible and highly localized micro-structured calibration phantoms within a transparent media. With the advent of the fs laser it is now possible to directly perform high-resolution micromachining and inscription in a wide range of planar and fibre devices. The fs laser inscription has been applied to some tough materials such as diamond, ceramic and silica. The interaction between ultra short laser pulse and a material is largely due to a non-linear process whereby multi-photon absorption is dominant. With a suitable numerical aperture-machining lens, with controlled machining translation speed and energy per pulse, it is possible to create extremely localized changes around a transparent laser focal spot within the material processed. In addition, by mounting the sample on a multi-axis stage it is possible to create complex three-dimensional structures inside it.

This thesis draws together research results on OCT phantoms and GPU data

processing to address two of the biggest issues in current OCT systems.

1.2 Thesis Structure

A brief overview of each chapter is given in the second part of this thesis, in which the contributions of co-authors work on all collaborative works (published and unpublished) for this chapter are briefly mentioned. Specific details of each co-author's contribution will be described throughout the thesis.

The thesis is structured according to the following scheme:

Chapter 1 outlines the background and motivation for the work described in this thesis and the structure of the this is given.

Chapter 2 outlines an overview of OCT systems, fs laser inscription and GPU (CUDA) processing. This includes a brief theoretical description of the OCT systems and fs lasers inscription as well as detailed description of NVIDIA CUDA GPUs software and hardware model.

Chapter 3 describes the requirements for an ideal *OCT-phantom* to serve assessment of the system resolution, sensitivity, distortion and other parameters.

Chapter 4 describes the three fs laser systems and the two OCT systems used during this study.

Chapter 5 presents the parameter space test (PST)¹ results. PST is a study that evaluates the combination of different inscription parameters such as inscription power, frequency, materials, inscription speed for producing a suitable OCT phantom. The investigation of suitability of fs lasers to inscribe structures for making an *OCT-phantom* is evaluated by studying the

¹PST is a study that evaluates the combination of different inscription parameters such as inscription energy, laser frequency, material, repetition rate and translation speed for producing a suitable OCT phantom.

PST using three different fs laser systems, each operating under different conditions. The PST samples were directly inscribed inside a variety of transparent materials (fused silica, borosilicate, doped-BK7 and Polymethyl methacrylate (PMMA)). The properties of the inscribed structures were characterized using an optical microscope (OM) and various OCT systems. In addition, cross sectional morphology of the inscribed structures were also studied for some of the samples with a scanning electron microscope (SEM) and a confocal microscope (CM) following cleaving and chemical etching processes. A high reproducibility of the inscribed structure was obtained. This shows high potential for producing multi-modality OCT calibration and characterization phantoms. This part of the work was carried in collaboration with *G.N. Smith, G.C.B. Lee, P.H. Tomlins and P.D. Woolliams* [1].

Chapter 6 describes the design, inscription and characterisation of the *OCT-phantom*. A single layered and a multi layered phantoms were fabricated inside a PMMA using a *Low-Rep 800 nm* system and a fused silica glass using a *Low-Rep 1026 nm* system respectively. These structures have been characterised using an OM and an OCT system. A multi layered phantom consist of four patterns, whilst a single layer phnatom contains one pattern. This part of the work was carried in collaboration with *G.C.B. Lee, P.H. Tomlins, P.D. Woolliams, M. Dubov and V. Mezentsev* [14].

Chapter 7 presents the development of quantitative characterisation methods of OCT performance employing *OCT-phantoms*. These methods use *OCT-phantoms* to experimentally measure the spatially variant resolution, lateral resolution, the sensitivity and sensitivity drop as well as distortion of an OCT (Michelson Diagnostics) system. More elaborate quantitative methods of distortion measurements and correction are under development, along with other quantitative methods such as scan linearity. This part of the work was carried in collaboration with *P.H. Tomlins, P.D. Woolliams, J.A.R. Williams, V.*

Mezentsev, M. Dubov and M. Baregheh [15].

Chapter 8 gives the results of OCT image enhancement such as deconvolution using the measured spatially varying PSF. Two different deconvolution routines, Wiener filter (WF) and Lucy-Richardson (LR) were used to deconvolve the OCT data using spatially varying PSFs. Here each B-scan is divided into tiles (sub-images) where the PSF is assumed to be spatially invariant, and each tile can be deconvolved using the corresponding average PSF of that region.

Chapter 9 describes GPU based data processing methods developed during this thesis. First section describes the standard OCT data processing methods. An approach to optimize OCT data processing using a GPU and parallel processing algorithms are presented. Standard data processing involves resampling the interference data from linear in wavelength to linear in wavenumber using interpolation methods, subtracting background, multiplying the results using a windowing function (spectral shaping), performing fast Fourier transformation (FFT) and finally obtaining absolute values and converting them to \log (or dB). Here two different interpolation methods (linear and third order Lagrange polynomial) as well as Hann windowing were implemented.

The final section of this chapter described GPU based deconvolution using spatially variant PSFs. Two different deconvolution routines using WF and LR have been implemented. This part of the work was carried in collaboration with *P.H. Tomlins, P.D. Woolliams, M. Dubov, J.A.R. Williams, V. Mezintaz [1].*

Chapter 10 briefly describes the development of an OCT system, which acquires and processes the data in real time (i.e. $\approx 92 kHz$). This part of the work was carried in collaboration with *P.H. Tomlins* at The Bart and School of Dentistry and Medicine, Queen Mary University of London.

CHAPTER 1. INTRODUCTION

Chapter 11 summarises the finding of the research reported in this thesis and also describes the layout of possible future direction for extending this work. Discussions include the significant of the reported work, and where appropriate possible directions of extending this are followed by the end of the thesis.

Several appendices are followed by the end of the thesis. An appendix (A) describes GPU-CPU based Gaussian fit approach used to estimate the PSF. Another appendix (B) details how the refractive index of the inscribed structures was estimated using a number of measurements including quantitative phase microscopy (QPM) images.

Chapter 2

Background

2.1 Optical Coherence Tomography

2.1.1 Introduction

OCT [7] is a non-invasive, non-destructive optical imaging technique [16], capable of producing three dimensional real-time volumetric tomography of a sample. The OCT approach builds on WLI and on OCDR, which was developed to find faults in optical fibres [2, 3]. The ability of OCT to provide the depth profile of the eye [4–7], and of other biological tissue was soon realised. The first successful demonstration of a depth profile along a single line or A-scan was demonstrated in early 1990s [17, 18]. By adding lateral scanning in a direction, transverse to the beam propagation, depth resolved 3D depth information can be captured [7].

OCT is based around an interferometric set-up as shown schematically in Figure 2.1 as a Michelson interferometer. The input optical field E_i , is split into a reference and a sample beam by an optical splitter. The light travels along the reference arm and sample arm and is reflected from the multiple layers, mirror E_r , and from the discontinuities of the refractive index changes within the sample E_s . These reflected fields are later re-combined at the splitter. Due to the low-coherence nature of the light source, interference between the sample E_s

CHAPTER 2. BACKGROUND

and reference E_r field occurs if the optical path length difference is less than the coherence length l_c of the light given in equation (2.1.1).

$$l_c = \frac{2\sqrt{\ln(2)}}{\Delta k} = \frac{2\ln(2)}{\pi} \frac{\lambda_0^2}{\Delta\lambda}, \quad (2.1.1)$$

where Δk is wavenumber spread, λ_0 is the central wavelength and $\Delta\lambda$ is the bandwidth.

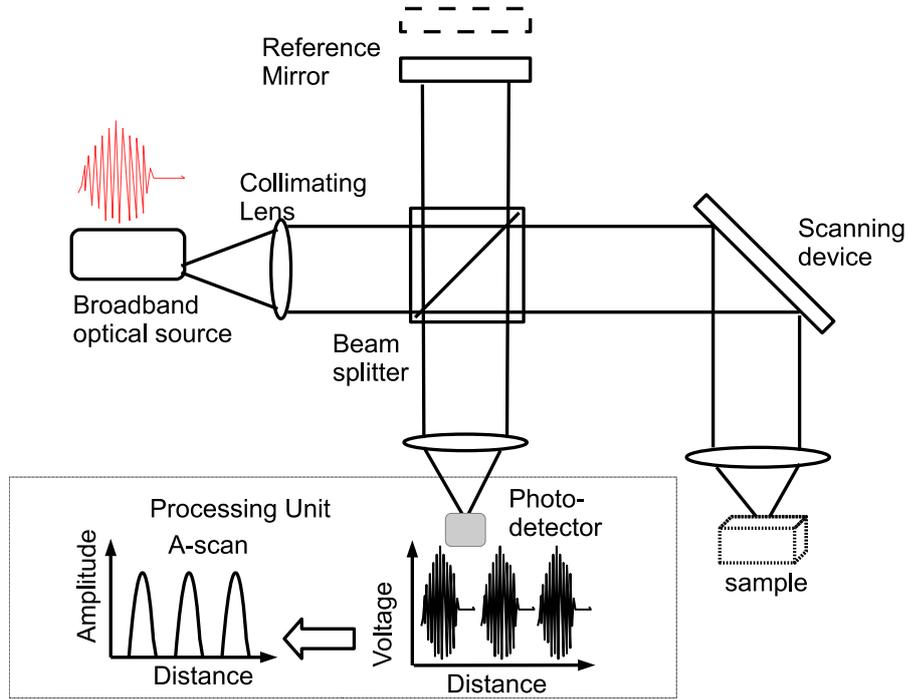


Figure 2.1: TD-OCT set-up.

Discontinuities in the refractive index of the sample appear as corresponding intensity peaks in the depth reflectivity profile (A-Scans). They can be measured either in time-domain or in the Fourier domain [16]. Multi-dimensional images are obtained by scanning the beam laterally and taking multiple depth-scans, orthogonal to the beam axis [16].

TD-OCT employs the principle of low coherence interferometry. As shown in Figure 2.1, TD-OCT uses an interferometer, a broadband (low coherence) optical source and a photo detector. Light from the broadband optical source is split

by optical splitter into sample and reference beams which are directed towards reference mirror and sample. A set of focusing optics and scanning mirrors are used to relay the object beam to and from object examined to the detection unit (i.e. photo detector). The reference beam is directed to a movable mirror called reference mirror.

Assuming the sample is composed of discrete reflective layers at different depths, each layer reflects the object beam with a delay corresponding to their depth. The layer satisfying the coherence gate condition, $OPD < L_c$, can be selected by scanning the reference mirror. Maximum interference is observed when $OPD = 0$. TD-OCT system produces the depth reflectivity profile by scanning the optical path difference (OPD), which is termed as an A-scan [19].

In the time domain (TD) approach, the depth profile is obtained by scanning the reference mirror placed in one arm of the Michelson interferometer, the reference arm. Interference is obtained when the path length of the light reflected from the refractive index discontinuities within the sample match the path length of the light in the reference arm. Reflectivity depth profiles can also be reconstructed in Fourier domain (FD) by fixing the reference path length (i.e. using a fixed reference mirror) and measuring the frequency components obtained by reading the optical spectrum of the reference output, by using a spectrometer [16]. This method is referred to as spectrometer based optical coherence tomography (SB-OCT). SB-OCT was demonstrated by Fercher et al [8]. Reflectivity depth profiles are reconstructed by Fourier transforming the spectrum read from the spectrometer. Swept-source optical coherence tomography (SS-OCT) is an alternative approach to SB-OCT, which uses a frequency tunable optical source (swept-source) and records the frequency components using a single detector [19].

Therefore, Spectral domain optical coherence tomography (SD-OCT) is classed into two types, as illustrated in Figures 2.2, depending on the source and detection methods used SB-OCT [8, 20–23] and SS-OCT (sometimes termed as optical frequency domain imaging (OFDI)) [24–26]. The major differences between SB-OCT and SS-OCT are the type of light source and the detection methods

CHAPTER 2. BACKGROUND

employed.

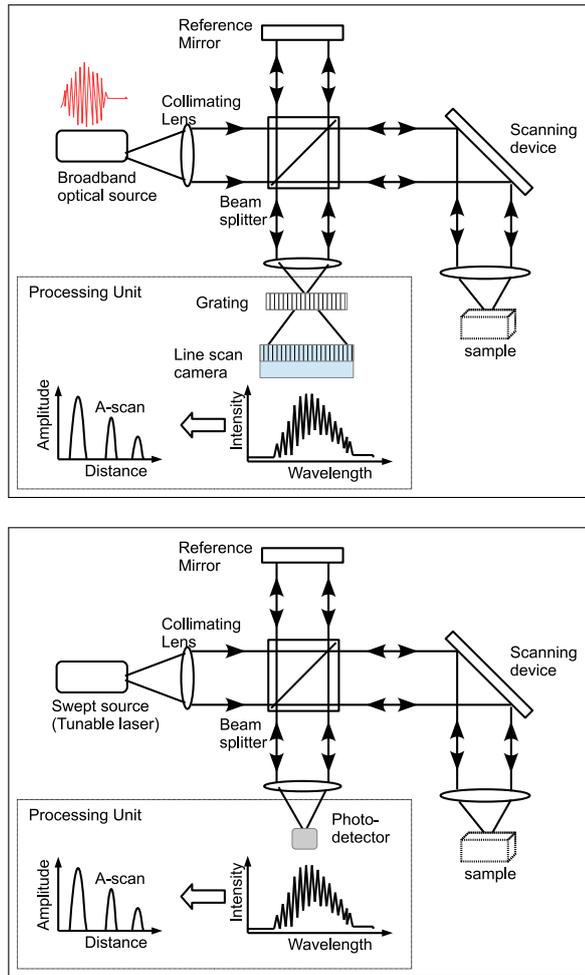


Figure 2.2: SD-OCT set-ups: SB-OCT (top) and SS-OCT (bottom).

SB-OCT uses a broad bandwidth low-coherence light source (i.e. super luminescent diode (SLED)) and a spectrometer employing a prism or diffraction grating and an array detector made of charge-coupled device (CCD) or complementary metal oxide semiconductor (CMOS) line scan cameras to disperse and detect the individual wave components [8, 19–23].

SS-OCT uses a narrow bandwidth, frequency swept (tunable) optical source and a single detector made of photo detector that records the intensity as a function of sweeping time while the laser is synchronously swept across its bandwidth [24–26].

In both cases a wave-number dependent intensity $I(k)$ is detected, proportional to the spectrum at the interferometer output. The depth $\sqrt{R_s(z_s)}$ profile information contained within this detected intensity can be determined by Fourier analysis.

In SD-OCT mechanical scanning used in TD-OCT system is replaced by other means. SB-OCT uses an array scan detector and diffraction grating whilst SS-OCT uses a frequency swept laser and a single detector [19].

SB-OCT and SS-OCT systems have sensitivity and acquisition speed advantages over conventional TD-OCT systems [16, 27]. Choma et al. have experimentally and theoretically showed that SD-OCT systems can have 30-40 dB more sensitivity over TD-OCTs systems [26]. The acquisition speed of the TD-OCT is relatively slow due to the need to mechanically vary the reference arm. Current SD-OCT systems can achieve acquisition rates of few hundreds to millions of A-scans per second [9–12].

2.1.2 Theory

The theoretical background behind SD-OCT is given in this section. TD-OCT is not described since SD-OCT dominates the current field/market, and all of the systems used in this study are spectral domain (SD) based.

The plane wave description of a complex electric field with source field amplitude $s(\omega)$, frequency ω and time t is given by:

$$E(\omega, t) = s(\omega)e^{-i(\omega t + kz)}, \quad (2.1.2)$$

The accumulation of phase in equation (2.1.2) is accounted for by the wave-number k and distance z . It can be assumed that the light source used is a polychromatic plane polarised complex electric field given by:

$$E_i = s(k, \omega)e^{-i(\omega t + kz)}. \quad (2.1.3)$$

CHAPTER 2. BACKGROUND

Assuming that an ideal beam splitter has a splitting ratio of 0.5 and that the reference mirror has a reflectivity of r_r , the electric field reflected from the reference reflector is given as [28]:

$$E_r = \frac{E_i}{\sqrt{2}} r_r e^{i2kz_r} \quad (2.1.4)$$

By assuming the sample is made up of N discrete layers, the sample response can be written as [28]:

$$r_s(z_s) = \sum_{n=1}^N r_n \delta(z_s - z_n), \quad (2.1.5)$$

where r_n and z_n characterise the discrete reflectivity and position of the reflectors within the sample respectively. The superposition of the total field reflected from the different refractive index discontinuities within the sample is [28]:

$$E_s = \frac{E_i}{\sqrt{2}} [r_s \otimes e^{i2kz_n}] \quad (2.1.6)$$

The detected field E_d can be expressed as the sum of the field returning from reference mirror, E_r and from the sample E_s [29]:

$$E_d = E_r + E_s. \quad (2.1.7)$$

E_r is made as a superposition of the fields reflected from the different refractive index discontinuities within the sample. If the sample is assumed to be made up of N discrete layers positioned at z_n distances measured from the beam splitter and has power reflectivity $R_n = |r_n|^2$, then the field (or sample response) from the n^{th} reflector can be given as

$$E_n = \frac{E_i}{\sqrt{2}} r_n e^{i2kz_n}. \quad (2.1.8)$$

CHAPTER 2. BACKGROUND

The overall signal from a sample with N discrete layers can be expressed as:

$$E_s = \sum_n E_n. \quad (2.1.9)$$

The intensity detected by an optical detector, relies on the square-law detection method, and is proportional to the time averaged product of the electric field and its complex conjugate:

$$I(\omega, t) = \langle E_d(\omega, t) E_d^*(\omega, t) \rangle \quad (2.1.10)$$

Substituting equation (2.1.9) into equation (2.1.7) and then equation (2.1.7) into equation (2.1.10) yields:

$$\begin{aligned} I_d &= \left\langle |E_r + \sum_n E_n|^2 \right\rangle \\ &= \left\langle (E_r + \sum_n E_n)(E_r + \sum_n E_n)^* \right\rangle. \end{aligned} \quad (2.1.11)$$

The power spectrum of the electric field is proportional to the intensity and the relationship can be given as:

$$P(\omega) = \frac{|E|^2}{2\eta_0} A, \quad (2.1.12)$$

where A and η_0 are the area of the detector and the impedance of free space respectively. The depth profile of the sample is determined by reconstructing the function $\sqrt{R_s}$ from the non-invasive interferometric measurement [28].

$$\begin{aligned} P(k) &= P_r(k) + \sum_n P_n \\ &\quad + 2\sqrt{P_r(k)} \sum_n \sqrt{P_n(k)} \cos(2k[z_r - z_n]) \\ &\quad + 2 \sum_{n,m \neq n} \sqrt{P_r(k) P_n(k)} \cos(2k[z_n - z_m]) \end{aligned} \quad (2.1.13)$$

CHAPTER 2. BACKGROUND

To simplify the approach, we can define the reference and sample arm reflectance as:

$$\begin{aligned} |r_r|^2 &= \frac{P_r(k)}{S(k)} \rightarrow P_r(k) = |r_r|^2 S(k), \\ |r_n|^2 &= \frac{P_n(k)}{S(k)} \rightarrow P_n(k) = |r_n|^2 S(k), \end{aligned} \quad (2.1.14)$$

where $S(k)$ is the power spectral density. Substituting P_r and P_n from equation (2.1.14) into equation (2.1.13) yields:

$$\begin{aligned} P(k) &= S(k)|r_r|^2 + S(k) \sum_n |r_n|^2 \\ &\quad + 2S(k)|r_r| \sum_n |r_n| \cos(2k[z_r - z_n]) \\ &\quad + 2S(k) \sum_{n,m \neq n} |r_r| |r_n| \cos(2k[z_n - z_m]) \end{aligned} \quad (2.1.15)$$

The Wiener-Khinchin theorem states that the auto-correlation function and the spectral density function are Fourier pairs [30]. Therefore, the spectral density function can be estimated by taking the Fourier transform of the autocorrelation function. The sample reflectivity profile is estimated from the inverse Fourier transform of the detected spectrum $I_d(k)$ or $P(k)$

$$F(z) = \mathcal{F}\{I_d\} \quad (2.1.16)$$

Using the convolution property of the Fourier transform $\mathcal{F}\{x(z) \otimes y(z)\} = \mathcal{F}\{X(k)\} \cdot \mathcal{F}\{Y(k)\}$ equation (2.1.16) can be written as:

$$F(z) = \mathcal{F}\{S(k)\} \otimes \mathcal{F}\{\mathbb{R}\}, \quad (2.1.17)$$

CHAPTER 2. BACKGROUND

where \mathbb{R} is

$$\begin{aligned}\mathbb{R} &= |r_r|^2 + \sum_n |r_n|^2 \\ &+ 2|r_r| \sum_n |r_n| \cos(2k[z_r - z_n]) \\ &+ 2 \sum_{n,m \neq n} |r_r| |r_n| \cos(2k[z_n - z_m])\end{aligned}\tag{2.1.18}$$

Using the Fourier transform pair of $\mathcal{F}\{\cos(kz)\} = \frac{1}{2}[\delta(z + z_0) + \delta(z - z_0)]$ the second term in equation (2.1.17) can be written as:

$$\begin{aligned}\mathcal{F}\{\mathbb{R}\} &= |r_r|^2 \delta(z) + \sum_n |r_n|^2 \delta(z) \\ &+ r_r \sum_n r_n^* \delta(z_r - z_n) + r_r^* \sum_n r_n \delta(z_r + z_n) \\ &+ \sum_{n,m \neq n} r_r^* r_n \delta(z_n - z_m) + \sum_{n,m \neq n} r_r r_n^* \delta(z_n + z_m)\end{aligned}\tag{2.1.19}$$

As noted in [31]: “the desired sample reflectivity profile $\sqrt{R_s(z_s)} = \sum_{n=1} \sqrt{R_{S_n}} \delta(z_S - z_{S_s})$ is embedded within the cross-correlation term of equation (2.1.19), surrounded by several compounding factors.” An A-scan of the OCT data can be obtained by making use of the shifting theorem of the convolution by delta function, and substituting the equation (2.1.19) into equation (2.1.17) [28].

$$\begin{aligned}F(z) &= \left(R_r + \sum_n R_n \right) \Gamma(z) \text{ "DC terms"} \\ &+ \sum_n \left[\sqrt{R_r R_n} \Gamma(z_r - z_n) + \sqrt{R_r R_n} \Gamma(z_r + z_n) \right] \text{ "Cross correlation terms"} \\ &+ \sum_{n,m \neq n} \left[\sqrt{R_r R_n} \Gamma(z_n - z_m) + \sqrt{r_r r_n} \Gamma(z_n + z_m) \right] \text{ "Auto-correlation terms"}\end{aligned}\tag{2.1.20}$$

In equation (2.1.21), the first and last term correspond to the DC terms, and auto-correlation terms. The second term corresponds to the cross-correlation components. The depth profile of the reflectivity of the sample is encoded within

the cross-correlation term.

2.2 Femtosecond Inscription

2.2.1 Introduction

The OCT calibration phantoms developed in this work rely on the unique properties of fs inscription. Highly precise deposition of energy in temporal and spatial position by fs lasers have great advantages over longer pulse lasers. These advantages lead to the use of fs lasers in a variety of applications such as laser surgery, photonics integrated optics, micro-fabrication and nanotechnology [32–36].

Internal material laser modification is a proven technique that enables one to fabricate complex three dimensional structures within the bulk of a transparent medium. This fabrication technique relies on the induction of the material modification at the focal region of the laser [37]. Highly focused fs laser pulses that are directed into the bulk of a transparent material can cause permanent structural changes inside the material confined within the focal volume of the laser beam. This occurs where the laser pulse intensity is high enough for multi-photon absorption (MPA) to occur leading to optical break down and micro-plasma formation. The formation of a micro-plasma causes localised modification of the structure of the material at the focal point by creating non-elastic thermochemical stress [38] or by forming colour centres [33]. This technique has the potential for writing complex three dimensional optical devices inside the bulk of a transparent material such as glass.

The fs induced bulk modifications in a variety of glasses (Ge-doped silica, borate, phosphate, fluorescence) and polymers PMMA have been demonstrated [34, 38–42]. A variety of devices and structures have been fabricated using this technique. This includes optical waveguides [33, 34, 38–48], optical amplifiers [45, 48, 49], beam splitters [38, 38, 45], directional couplers [34, 34, 46], three di-

mensional data storage [50], birefringent transmission grating [51], transmission grating [52] and long period fibre grating [51].

The magnitude of the modified refractive index has been estimated indirectly from numerical aperture (NA) measurements and far field pattern analysis [40], refractive near field techniques [34, 40] and measured using a microelipsometer [33]. Furthermore, modified regions have also been chemically etched to form microfluidic channels [36, 50, 53, 54].

Highly focused fs pulses with moderate energy can have extremely high peak powers and peak intensities [47]. The laser intensity at the focal point can reach around 10^{13} to 10^{14} W/cm^3 [55]. When the material interacts with a highly focused pulse, the material response becomes highly non-linear [47]. This leads to a metal-like behaviour of the dielectric [56–58] where for example the index of refraction becomes intensity dependent and near infra-red (NIR) light is absorbed by the transparent material, generating free electrons [47]. The high intensity at the focal point leads to significant ionization and to formation of a plasma. The presence of plasma further causes MPA and plasma defocusing [59].

When a fs pulse propagates thorough a transparent material it can self-transform into a broadband pulse [60]. A phenomenon known as super-continuum generation occurs during the filamentation process, which is induced by the dynamics balance between self focusing by non-linear Kerr effects and self defocusing by plasma formed during non-linear ionisation process [55].

2.2.2 Theory

The underlying mechanisms of matter with ultra-short pulses are still unclear. Generally the mechanism is divided into two parts: energy absorption by non-linear ionisation, that is treated in a similar way to laser ablation modelling, and dissipation of the absorbed energy [61, 62]. The underlying mechanism of energy absorption in ablation has been studied broadly [39, 53, 63–65]. It is known that the interaction, at the focal region of the beam, between the ultra short laser

pulse and the material is a non-linear MPA. This is because most glasses do not absorb NIR light through a linear single photon process. Upon absorption of laser energy, free electrons are produced by multi-photon ionisation (MPI) of the glass matrix. This is a deterministic process which eliminates the need for presence of impurities or other defects in glass to produce the seed electrons.

In normal circumstances, dielectric materials are transparent to NIR light¹. However, non-linear absorption is realised when the intensity of the incident light is high enough. The short pulse duration ensures that there is no time for coupling of electron energy to the surrounding lattice resulting in a localised energy deposition and plasma formation [53]. NIR light does not have enough energy to free a valance electron in dielectrics, because the minimum energy required to break the electron lattice bond is much greater than that available in a single photon.

Formation of plasma occurs when the absorbed energy reaches a critical threshold [61]. Maxwell's equations govern the propagation of electromagnetic waves in transparent materials. The wave equation of an electric field \vec{E} inside a dielectric, considering only the dipole response of the material, can be expressed as [39, 66]

$$\nabla \times \nabla \times \vec{E} + \mu_0 \epsilon \frac{\partial^2 \vec{E}}{\partial t^2} = -\mu \frac{\partial^2 \vec{P}}{\partial t^2}, \quad (2.2.1)$$

where \vec{P} is induced polarization. Polarization in the linear regime is related to the electric field by [39]:

$$P(r, t) = \epsilon \chi^{\{1\}} E(r, t), \quad (2.2.2)$$

where χ is the linear susceptibility tensor. Here the material response is assumed to be instantaneous. At higher intensities, the linear polarization used in equation (2.2.2) is no longer valid, and should be replaced by the non-linear term

¹The binding energy of fused silica is $9eV$.

CHAPTER 2. BACKGROUND

for the polarization given as [39]:

$$\begin{aligned} P &= \epsilon (\chi^{\{1\}} E + \chi^{\{2\}} EE + \chi^{\{3\}} EEE + \dots) \\ &= P^{\{1\}} + P^{\{2\}} + P^{\{3\}} + \dots, \end{aligned} \quad (2.2.3)$$

where χ^n is the nth order susceptibility and P^n is the nth order polarisation. Here explicit time and spatial indices are dropped.

Amorphous materials are isotropic and have inversion symmetry, hence equation (2.2.3) can be simplified to [39]:

$$P = \epsilon \left(\chi^{\{1\}} + \frac{3}{4} \chi^{\{3\}} |E|^2 \right) E. \quad (2.2.4)$$

This gives the index of refraction as [39]:

$$n = \left(1 + \chi^{\{1\}} + \frac{3}{4} \chi^{\{3\}} |E|^2 \right)^{1/2} \quad (2.2.5)$$

$$\approx n_0 \left(1 + \frac{3 \chi^{\{3\}} |E|^2}{8 \chi^{\{1\}} + 1} \right), \quad (2.2.6)$$

where the linear refractive index $n_0 = \sqrt{1 + \chi^{\{1\}}}$.

Assuming the nonlinear term is much smaller than the linear susceptibility and expressing the electric field strength in terms of the laser intensity I , index of refraction can be expressed as:

$$n = n_0 + n_2 I, \quad (2.2.7)$$

and the nonlinear refractive index is given as:

$$n_2 = \frac{3 \chi^{\{3\}} |E|^2}{8 \chi^{\{1\}} + 1}. \quad (2.2.8)$$

The self focusing and self phase modulation effects are results of the nonlinear refractive index [39].

Because of the large band gap, the energy available in single photon is insufficient to break the ground energy (band gap energy). Structural modification and optical break down of the transparent (dielectric) material occurs by non-linear absorption of the laser light. Multiple photons must be simultaneously absorbed, where the combined energy must be greater than the band gap, in order to excite an electron. Non-linear photo-ionisation (NLPI) is generally grouped into MPI or tunnelling regimes depending on the intensity and frequency of the laser field. Although the conceptual pictures of these two regimes greatly differ, they can be describes by a Keldysh framework.

Tunnelling ionization becomes dominant when the laser has a strong field and a low frequency. The strong laser field suppresses the Coulomb well that binds the valance electrons, thus electrons can tunnel through the short barriers². At higher frequencies MPI dominates, in which the electrons are excited by simultaneous absorption of multiple photons. The transition point from MPI to Tunnelling is described by the Keldysh parameter given as [39, 67]:

$$\gamma = \left(\frac{\omega m c n \epsilon_0 E_g}{e I} \right)^{1/2}, \quad (2.2.9)$$

where I is the intensity of the laser, e the electron charge, n the refractive index of the material (i.e. glass), c the speed of light, m the effective mass of electrons and ω and E_g are the angular frequency and the electric field of the laser respectively. MPI dominates when γ is much larger than 1, while tunnelling is realised when γ is smaller than 1. In the intermediate regime i.e when $\gamma \approx 1$ the photo ionization is a mixture between tunnelling and MPI [62].

Even though MPI is a common process, the lack of intensity of the incident light means this process only contribute to a fraction of a single photon absorption [61]. In order for MPI to become a dominant process, the following two conditions need to be satisfied:

1. The band gap of the material must be greater than the energy of a single

²DEFINE SHORT BARRIERS.

CHAPTER 2. BACKGROUND

incident photon,

2. The intensity of the light must be high enough so multiple photons can be absorbed simultaneously.

The rate of multi photon absorption can be described by the equations [61]

$$\frac{dN}{dt} = \sum_m \gamma_m I^m, \quad (2.2.10)$$

where $\frac{dN}{dt}$ refers to the number of photons excited per unit time, γ_m is the frequency dependent absorption coefficient for the m^{th} order absorption and the intensity of the incident laser field is denoted by I [61].

The photon intensity increases by both single and multiple photon ionisations. Therefore, linear ionization must be suppressed for MPI to dominate. This can be achieved by adjusting the frequency of the individual photons incident on the material.

The rate of MPI strongly depends on the intensity and is described as

$$P(I)_{MPI} = \sigma_k I^k, \quad (2.2.11)$$

where σ_k is the MPI coefficients [39, 63] and k is the smallest number of photons required to satisfy the condition $k\hbar\omega \geq E_g$ [67].

In the presence of the laser field, the electron density N in the conduction band grows as ηN , where η is the avalanche ionisation rate. The initial seed electrons required for the avalanche ionisation are provided by mechanisms such as MPI, tunnelling and thermal excitation. The formation of plasma, by nonlinear ionisation, absorbs further laser energy. The absorption rate is efficient when the plasma frequency ω_p given by [39]:

$$\omega_p = \left(\frac{Ne^2}{\epsilon m} \right), \quad (2.2.12)$$

reaches the laser frequency. The plasma abortion frequency is given as [39, 68].:

$$\kappa = \frac{\omega_p^2 \tau}{c(1 + \omega^2 \tau^2)}. \quad (2.2.13)$$

The spatial distribution in the beam causes a spatial distribution in the plasma. The laser has a Gaussian intensity profile, thus the photon density is highest at the centre of the focussed beam and decrease radially outward. This acts as a negative lens that defocuses the pulse, which in turn halts the catastrophic collapse predicted by self focusing.

The non-linear absorption of photon occurs in the following order: first non-linear photo ionisation, followed by avalanche photo ionisation and finally plasma formation occurs [62]. The non-linear ionisation can deposit energy in a region smaller than the wavelength of the incident laser light [37]. This small energy is insufficient to damage the material but sufficient enough to make structural changes [33, 69–72]. The inscription process is dependent on the applied field intensity and only starts if the intensity exceeds a critical value I_{th} [73].

Photo ionization refers to the process of transferring an electron from the valance band to the conduction band. In transparent materials, this band gap is quite large, thus the energy carried by a single photon is insufficient to cause this transition. Electrons can absorb multiple photons and be excited from the valance to the conduction band.

For incident light with lower energy and high intensity tunnelling, ionisation becomes the dominant process. In the tunnelling ionisation regime, the strong field suppresses the Coulomb potential well and electrons become free by tunnelling through the potential barrier [39].

The deposition of laser energy into the material on a time scale much shorter than the thermal diffusion causes ablation on the surface or structural modification in the bulk of the material [62]. Formation of plasma strengthens the absorption of the laser energy and can modify the real part of the refractive index according

to [62]:

$$n = n_0 - \frac{N}{2n_0 N_{cr}}. \quad (2.2.14)$$

Avalanche ionization involves free carrier absorption followed by impact ionization [39, 62]. Seed electrons required for the avalanche ionization are provided by one or all of the following processes; multi photon and tunnelling ionization, thermally excited carriers, easily ionized impurities or defects [39, 74]. These free electrons absorb photon energy via the inverse Bremsstrahlung process (consecutive linear absorption of several electrons) and move to higher energy state in the conduction band [39, 62]. An excited electron ionises another electron from the valence band, when its energy exceeds the minimum conduction band by more than the band gap energy [62, 75]. This causes two excited electrons in at the conduction band. Avalanche ionization dramatically increase the concentration of conduction band electrons [61, 76, 77] and the growth can be expressed as [62]:

$$\left(\frac{dN_e}{dt} \right)_{avalanche} = \alpha I N_e, \quad (2.2.15)$$

where α is the avalanche ionisation coefficient. The density of the electrons in the conduction band grows until the plasma frequency approaches the frequency of the incident laser called critical plasma frequency.

The formation of plasma can increase the absorption of laser energy, at the same time negative refractive index of the free electrons in the plasma can have a defocusing effect on the incident laser pulse. The spatial distribution of electrons is higher at the centre and decreases radially outward due to the intensity profile of the beam [39]. This counteracts the effect of self focusing and stops the catastrophic collapse of the beam to singularity. Non-linear index is balanced by the negative refractive index contribution of the plasma, when plasma density reaches $10^{17} - 10^{18} \text{ cm}^{-3}$ [39, 75].

CHAPTER 2. BACKGROUND

Self focusing arises from the intensity dependence of the refractive index [39] variation caused by variation in the intensity of the incident pulse [62]. The refractive index at the centre of beam is greater than that of the wings. This distribution in the refractive index acts as a lens and focus the incident beam [39, 62]. Self focusing remove the restriction of linear focusing of the laser beam [37]. As refractive index variations depend on the intensity of the incident laser pulse, the strength of the self focusing lens depends on the peak power of the laser pulse [39, 78, 79]. Self focusing increase with increasing peak laser power, until a critical point where it balances diffraction and filament is formed. The critical power P_{cr} is expressed as [39, 62, 79]:

$$P_{cr} = \frac{\pi(0.61)^2 \lambda_0^2}{8n_0 n_2}, \quad (2.2.16)$$

where λ_0 is the vacuum wavelength of the laser radiation.

Only an incident beam having a Towne mode profile (a special non-linear profile) will experience self-focusing at $P_{in} = P_{cr}$, and for all other beam profiles, the critical power required for self-focusing is much greater than P_{cr} [37]. The power required for self focusing of a Gaussian beam is very close to P_{cr} [37]. Catastrophic collapse of the beam to a singularity will occur when the peak power of the incident laser pulse exceeds critical power for self focusing [39, 62, 79]. However, in practice, catastrophic collapse is prevented by free electron plasma created by the non-linear ionization. Plasma acts as a negative refractive index change that prevents the self focusing [62, 80]. As self focusing increases, the intensity increases with it and become sufficient enough to non-linearly ionize the materials. The formation of plasma, having negative refractive index, counteracts the positive index of the self focusing and halts the self focusing further [39, 80].

Laser pulses have spatial and temporal intensity distributions. The spatial intensity distribution leads to a variation in the refractive index change. Self focusing arises from the intensity dependent refractive index. Intensity of a Gaussian beam is higher at the centre, thus gives a higher refractive index at the centre

of the pulse and decreases radially outward. This spatial distribution in refractive index acts as a lens and focuses the beam. Even though the refractive index variation is related to the pulse intensity variation, the strength of self focusing depends only on the peak power of the pulse [39, 78]. Therefore self focusing increases with increasing pulse power. Dynamic balancing between self focusing and diffraction occurs at the critical power and a filament is formed. The beam catastrophically collapses to a singularity if the power exceeds this critical power [39, 78]. The catastrophic collapse due to self focusing is prevented by other mechanisms such as plasma formation and self phase modulation.

Self phase modulation is the temporal analogue of self focusing. Temporal intensity variations in the pulse give rise to a time depending refractive index. This in turn produces time dependent phase shifts of the pulse.

Since fs lasers can be used to modify the refractive index of transparent materials we can use this to inscribe 3D optical devices such as waveguides. Waveguides can be written in either longitudinal or perpendicular direction to the propagating beam. Since the Gaussian beam has a radial symmetry, longitudinal writing generates a radially symmetric structure. However, the drawback of this method is that the structures should be written in the longitudinal direction (i.e. parallel to beam axis). Hence the length of the structure is restrained to the working distance of the objective [61].

There is no constraint imposed by the objective when the structures are inscribed in the transverse (perpendicular) direction. In addition, this also allows to write structures with more flexible shapes [61]. However, the symmetry of the structure strongly depends on the focusing conditions [45]. The lines written in the transverse direction have an elongated profile due the elongation of the Gaussian beam focus [61]. This elongation can be minimised by using a high NA objective [61]. The drawback of a high NA objective is that depth at which the structures can be inscribed within the bulk of the material is reduced [61]. Alternatively, elongation in a single transverse direction can be minimised by employing a cylindrical lens or a slit to create an elliptical beam [62, 81]. The consequence

of this is that the structure shapes dependence on the writing direction and only straight lines are possible [61]. Structures written perpendicular to the laser axis have elongated structures for all pulse durations [62]. The elongation varies with pulse energy. Elongation in the axial direction (i.e. parallel to the beam) increases dramatically with increased pulse energy compared to that of transverse direction [82]. It has been observed that the elongation in the axial direction is always greater than the depth of focus. This indicates the presence of self focusing and filamentation [56].

Induced material modification or optical break down strongly depends on the pulse duration [83]. The fs laser has several advantages over longer pulse lasers. The peak intensity at the pulse wing is much smaller than the threshold energy, thus produces negligible collateral damage surrounding the focal region [64].

2.3 Parallel Processing

2.3.1 Introduction

The acquisition speed of SD-OCT dramatically increases every year. Current SD-OCT systems, based on line scan cameras have been demonstrated to achieve up to $312,500 \text{ lines/sec}$ [9], while SS-OCT systems have achieved over multi-MHz using Fourier domain mode locked (FDML) systems [84]. In 2005 Huber et al., demonstrated a 20 kHz SS-OCT [85], Potsaid et al., later demonstrated an SD-OCT based system that is capable of recording A-scan at rates from 70 kHz to 312.5 kHz [9] and Gora et al., demonstrated an SS-OCT system which achieved an acquisition speed of 200,000 A-scans per second [86].

The high acquisition SD-OCT systems are capable of producing dynamically resolved (3D space + time), high resolution images. This allows recording of dynamic processes such as pupillary reaction to light stimulus [13, 86] and embryonic heart beating [87–89] in three dimensions [90]. The high acquisition is an attractive attribute for micro-surgical guidance and interventions [90].

CHAPTER 2. BACKGROUND

One of the drawbacks of the SD-OCT systems is that they require significantly more data processing to handle the large amount of information recorded. Historically this has incurred a significant time overhead. Traditional single central processing unit (CPU) based processing cannot cope with these high throughput system and so the processing becomes a bottleneck. Generally the data acquired from these systems are saved then processed and rendered at a later time.

Processing of OCT data is a computationally intensive task. Increased data acquisition means, increased data processing. However, data processing and rendering of OCT has not kept up with the advancement of the acquisition rate. Therefore, processing large amount of the volumetric data acquired by OCT systems becomes a processing bottleneck that severely limits the performance of the OCT for real time measurements. Many of the current OCT systems capture and save the raw data and then process and renders at a later time [90]. Due to the limited processing, current processing is only able to process and render two dimension (2D) (B-scan) in real time or a subset of the actual data instead of the full volume. Processing of all the volume/data in real time is important to clinical applications such as endoscopy and ophthalmology [90]. Real-time acquisition, processing and rendering of the OCT data must be realised to fully utilize the OCT.

Resampling by interpolation is a computationally expensive calculation in the SD-OCT data processing, if resampling is not needed then the total required processing time reduces dramatically. However, linear k detectors are not absolutely linear for wide spectral regions. Therefore, resampling is still necessary to achieve high axial resolutions [91].

Several solutions such as field programmable gate arrays (FPGAs), multi-CPU and GPU have been proposed to overcome the processing bottleneck. Generally, increased processing is achieved by FPGAs and digital signal processors [92–96]. This additional hardware, increases the complexity and cost of the systems. Parallel CPU based processing has also been demonstrated [97] that achieves a processing rate of 80,000 *lines/sec* on a non-linear k system [97] and

207,000 *lines/sec* on a linear k system where each A-scan had 1024 pixels [98]. More recently CPU-GPU hybrids architectures have been used to enhance the processing rate of the OCT systems. This method has been applied to both linear in k [99] and non-linear system [90, 100–102].

Several solutions based on FPGAs [103], signal processors, multi threaded CPUs [97, 104] and GPUs [90, 99–101] have been proposed to solve the processing bottleneck. In addition to accelerating the data processing, reduced data processing is realised by means of linear in wavenumber (k) swept light sources [104, 105] and linear in k detectors [99], which avoids the unnecessary resampling from linear in wavelength to k space. GPUs require parallel programming algorithms to exploit their performance power, application programming interfaces (APIs) are available such as CUDA and OpenCL (OpenCL), which can be used to program GPUs for general purpose computations.

GPUs have been used elsewhere in scientific computing to accelerate processing times. Since OCT data is highly parallel, it is therefore ideal to accelerate the processing using GPUs. GPUs are highly suitable for high parallelism and memory bandwidth and their processing for OCT was first realised in 2009 [99]. Watanabe and Itagaki [99] used a GPU (NVIDIA GeForce GTX 280) with a linear-in-wave-number spectrometer to achieve a processing speed of 55,800 *lines/s*. Subsequently, Zhang and Kang [90] used linear-spline interpolation and a GPU (NVIDIA Quadro FX5800) to accelerate the processing and rendering of OCT data from a non-linear k -space spectrometer. They reached a processing rate of 680,000 *lines/s* for overall volume processing. However, for smaller A-scan numbers the processing bandwidth decreased to just greater than 200,000 *lines/s* for the range of 1,000 to 10,000 A-scans.

Van der Jeught et al. [100] compared nearest-neighbour, linear and cubic-spline interpolation to resample algorithms running on a GPU (Geforce 9800GT) and achieved a processing speed of 25,600 *lines/sec*.

2.3.2 CUDA

The video gaming market is the driving force behind the development of GPUs. Due to different requirements, they have a different architecture compared to their counterpart CPU. They were designed as massive parallel, high throughput systems. General-purpose computation on graphics processing units is well established in scientific community. Highly parallel data such as OCT data can be processed with speed many times faster than that is possible with conventional CPUs.

Even though the immense power of GPU has been realised for general purpose scientific computing, exploiting the power of the GPU hardware is a prodigious task which uses low level programming such as OpenGL (OpenGL). Programming close to the machine level is required, that is only achievable by an elite group of experts. However, in the last few years, high-level APIs have been developed, which allow programmers to utilise the GPU hardware without in-depth knowledge of the underlying hardware (i.e similar to CPU programming). One of such high level API is the NVIDIA's CUDA.

CUDA is a programming paradigm provided by NVIDIA, comprised of tools and libraries to exploit parallelism within the NVIDIA GPUs [106]. This allows the programmers to express the parallelism explicitly for NVIDIA's GeForce G8 series GPU or one of its successor architectures and write scalable parallel code to leverage the parallel processing power of NVIDIA's GPUs for scientific computing. The CUDA framework is comprised of a hardware driver, API and its runtime, and higher-level mathematical libraries and a compiler [106, 107]. It can be easily integrated into existing C/C++, and other programmes. As the CUDA is an extension to C, learning curve is small compared with other GPU programming in the market. However, the only impediment of CUDA is, it does not support other vendors to date.

Here description of both hardware and software architecture of CUDA is given. Later performance of the system and comparison between CPU and CUDA pro-

programming concepts are given. Brief details of other GPU programming models are also presented by the end of this chapter.

CUDA Hardware Architecture

The GPUs act as a co-processor to the CPU, where data independent compute intensive portion of the calculation can be off loaded for processing. Current GPU architectures have several noticeable features that are important for general-purpose programming. Host interface - allows programmers to easily interact with GPU, copy data from GPU main memory to GPU memory. Texture/processor cluster (TPC) - is the basic building block of CUDA hardware architecture. Each TPC contains two streaming multiprocessors (SMs), and a Texture Unit that is shared by both SMs [108].

The GPUs are constructed by aggregating multiple SMs, each in turn comprising multiple scalar processors (Arithmetic logical unit (ALU)). Each SM contains several (i.e 8 in G8 series) ALUs, a special function unit (SFU) for performing interpolation and approximate evaluation of trigonometric functions, a multi threaded instruction fetch and issue unit, thread caches (registers) and shared memory. A system known as compute work distribution (CWD) manages the resources and the execution of threads by assigning a group of threads to a SM and performs context switching [106].

CUDA allows scatter and gather (vectored input/output (I/O)) operations from device memory hence data can be read from and written to any location in the global memory as flexible as CPU. Reading data from global memory is a long latency process that requires between 400 and 600 clock cycles. On chip shared memory is used to avoid frequent global memory access. Reading data from shared memory requires only 4 clock cycles. Therefore data can be read from and written back to arbitrary memory spaces [97, 106]. In order to achieve fast thread communication, CUDA features a parallel data cache or on-chip shared memory with very fast read and write support [106].

CHAPTER 2. BACKGROUND

Each processor has access to several memory spaces. All processors have access to 32 bit registers, private to that thread that runs on it. All the processors in the SMs have access to the shared memory on that SM, that can be used to synchronisation and communication between the threads within the SM. If bank conflicts can be avoided, then shared memory can be accessed as fast as registers. All the SMs on the card has access to global, constant and texture memory spaces. The type of the memory, location on the GPU, available size and scope of the variables are summarised in the Table 2.1.

Table 2.1: GPU hardware memory hierarchy. R/W: Read and write; R: read only.

Memory	Location	Access	Scope
Register	on-chip	R/W	one thread
Local	off-chip	R/W	on thread
shared	on-chip	R/W	all threads in a block
Global	off-ship	R/W	all threads + host
Constant	off-chip	R	all threads + host
Texture	off-chip	R*	all threads + host

The main GPU used for the research is Tesla C1060 GPU. This consists in a total of 240 ALUs (processors) and 30 double precision units (DPUs), arranged in 30 SMs. Each SM has 16 *kB* shared memory, 16 *kB* Constant memory, 2^{32} 32bit registers. Additionally, the device has a 4 *GB* global (dynamic random access memory (DRAM)) memory that is accessible by all of the threads.

CUDA Software Architecture

Threads are the lowest construct in the CUDA software model that are then grouped into 32 threads called warps which are further divided into half warps (i.e. 16 threads). Threads within each half warp execute the same instruction simultaneously on different data, similar to single instruction, multiple data (SIMD) fashion. If there is a difference in the execution of the threads, then only that particular half warp is serialised.

Threads are grouped into up to three dimensional thread blocks. Threads within any block can be identified by an internal index. The programmer can allo-

CHAPTER 2. BACKGROUND

cate the number of threads arbitrarily during the runtime. The maximum number of allowable threads per SM and per block depends on the computational compatibility of the hardware being used. For example, the card Tesla C1060 allows a maximum of 512 threads per block whilst maximum of 1024 threads can be active on each SM.

Blocks are arranged into up to two dimensional blocks of grids. Blocks are scheduled to one of the available SM by a thread scheduler and expected to run independently, executing the same code, on different part of the data. This procedure is similar to single programme, multiple data (SPMD) concept, where parallel code (same instruction set) is executed at different notes on different subsets of the data. This is a fundamental requirement of the CUDA model. The grid is the highest level of construct in CUDA software model. A grid contains all the threads available to execute the instruction specified by the kernel.

When the warps have to wait for a long latency operation (e.g. reading data from global memory), execution is stalled. When this occurs, context switch switches to another warp that becomes ready to be executed. This hides the long latency process and avoid the processors going to an idle state. This switching of blocks hides the long latency processes, given that there are enough blocks available to be executed by an SM. Therefore a minimum of two blocks must be available in order to hide the latency.

Intra communication between the threads within a block is possible via shared memory and they are intrinsically synchronised. Inter communication of threads from different blocks are not possible.

CUDA employs a new architecture called single instruction, multiple threads (SIMT) model, which is a combination of SIMD where each threads execute the same instruction on different data and SPMD where autonomous blocks execute same instruction independently on a subset of data. However, SIMT allows programmers to specify execution and branching behaviour of a single thread. If there is a divergence in the execution of a warp, hardware automatically serialise execution of that warp. This gives the programmers freedom to write thread level

independent parallel codes [106].

Each thread within a block is executed in SIMD fashion. The data are subdivided into multiple subsections, each is processed by different blocks simultaneously in SPMD fashion. This means that the intra-communication between the threads within a block is possible via shared memory, but communication between different blocks is not intrinsic.

Threads within warps are executed simultaneously or remain idle when branching occurs [108]. The warps active on each SM are scheduled by special units in SMs in such a way that, without any overhead, several warps execute concurrently by interleaving their instruction [108].

A kernel is a function that is invoked by CPU and executed on the GPU in SIMT fashion to process the parallel portion of the code. The kernel called by CPU requires configuration parameters, which specify how to allocate the threads to blocks and blocks to grid. This is one of the important aspects of achieving performance gains. Allocation of number of threads per blocks (active thread per SM) is limited by many factors. Maximum size of the index of grid and block in any dimension is 2^{16} .

The kernel invocation by the CPU does not block the host and the device and host are not implicitly synchronised. When the CPU invokes multiple kernels, the scheduler schedule them without overhead so they will be executed in the order they have been called by CPU. Only a single kernel will run at a time in devices with compute comparability 2.0 or less. Therefore, when multiple kernels are called by CPU, kernels must wait until the previous one is complete. Devices with compute computability greater than 2.0 can run multiple kernels thus, kernel overhead is possible. Device memory access is blocked by default when a kernel is executed on the device (i.e. memory is implicitly synchronised) [106].

The blocks are scheduled to one of the SM during the runtime. Multiple blocks can be scheduled to the same SM and executed concurrently. Allocating multiple blocks to the same processor means, that resources are split among the active blocks. The number of active blocks scheduled on a SM depends on the number

of threads in the block and the amount of available shared memory and registers. This is also limited by the maximum allowed by software model, which for example on Tesla 1060 the number of threads is limited to a maximum 768 threads per block or 1024 active threads.

The software limits the maximum of threads per block to 768 and the number of threads that can be active per SM to 1024. For example, one block with 768 threads or two blocks with 512 threads each can be active per SM. Shared memory of 16 *kB* and registers in each SM are equally shared between the active blocks (and active threads). For example when a block is equipped with 768 threads (maximum number per block) only one block will be active per SM, hence the total shared memory can be used to store the data of this block and the registers are split up between the threads (i.e. ≈ 21 registers per thread). On the other hand if two blocks of 512 thread are active (total of 1024 threads), then each block has access to 8 *kB* of shared memory and 16 registers per threads.

During the execution, threads are divided into SIMD groups called warps. The current warp size is 32, and there are eight processors within each multiprocessor, which can concurrently execute the kernel in SIMD fashion by scheduling them on the processors of a multiprocessor. Consequently, execution of a single warp requires a minimum of four clock cycles. When the blocks are ready to be executed, threads within the blocks are divided into half warps and scheduled by the special unit (thread scheduler), which schedules the warps without overhead. Each thread within the warp reads/fetches the data according to its instruction and thread index.

The number of active blocks per SM falls by block granularity, and a minimum of one block is required to run the programme successfully, else the kernel will fail to execute. The index of any thread within the grid can be unambiguously identified using the block index and thread index inside the block. Device can only access one of the memory available in the device, therefore data need to be transferred to the GPU prior to any processing.

CUDA memory (software) hierarchy model

The CUDA software memory hierarchy is slightly different from that of the hardware memory hierarchy. All threads in a grid have access to global memory and read-only access to constant memory. Shared memory is defined in block scope, which means that all threads reside within a block that can access this memory space. However, accessing shared memory from another block is not allowed.

The global memory is accessible by all the threads specified within a grid. Additionally, the host can access the global memory using run the time library. Variables declared in the global memory space have a lifetime of the application, the variables declared in the shared memory and registers have the life time of the block. Current CPUs supports scatter and gather memory access with byte-wise memory addressing [109]. Non coalesce memory access may lead to serialized reading or writing, thus dramatically reducing the performance [109].

Accessing the global memory is one of the longest latency processes in CUDA programming model [110]. Access to the global memory may take as 400-600 clock cycles. Access to (coalesced) shared memory takes about 4 clock cycles. Therefore, avoiding direct access of data to the global memory should be avoided. Data should be transferred to the shared memory when multiple access is required. This improves the performance of the application.

Coalescing the (global) memory access reduces the number of access request. Coalesced memory has consecutive addresses; therefore they can be read by a single access. If the data access is not coalesced, then the access will be serialised. Coalesced memory access exploits the full potential of the memory band width. Thus, data required for the entire block can be transferred in a single access rather than one memory access per thread in the block, potentially saving hundreds of cycles.

A program must hide the slow global memory access and other long latency process by ensuring there is enough work/wraps such that during the waiting time, current warps can be switched to one that is ready for execution. Coalescing

CHAPTER 2. BACKGROUND

the global memory access is the most important aspect of performance. Global memory access by the threads in a block is split into half warps. Each warp can be accessed by as few as one transaction, when certain memory alignment criteria, named coalescing are met.

Unlike reading from device memory, writing to global memory takes fewer clock cycles that can be considered as a fire-and-forget instruction. Processors send the instructions to special memory writing unit and run the rest of the warps. This means writing intermediate results to global memory reduces the performance. For example, let us assume that for a certain computation three steps are required and, at each step, the data is read from and written back to the global memory. This means that threads must be synchronised between steps to make sure that data is completely written to the global memory prior to the next step. This increases the burden.

Each SM has a pool of small shared memory space, which is divided among the number of blocks active on the SM. Any variable in shared memory has a life time of the block. The shared memory has very low access latency and extremely high memory throughput. If there is no bank conflict, shared memory throughput is almost comparable to the registry throughput [109]. To gain maximum performance benefit of using shared memory, the programmer has to address the access conflicts and alignment of data structures.

The shared memory latency is nearly 100 times smaller than the global memory latency, providing that there is no bank conflict. The shared memory is divided into banks, 32 bit words, to provide high bandwidth memory access. Accessing n distinguish addresses can be accessed simultaneously without conflicts, which results in highly effective bandwidth [111]. However, bank conflicts arise if multiple addresses request access to the same memory bank, in which case the access is serialised. Hardware splits the access request into separate multiple conflicts free requests. This reduce the effective memory access bandwidth by a factor given by the number of separate access.

Texture and constant memories are specialised memories, which reside in

the device/global memory and are cached. All the threads and the CPU using runtime library has access to both of these memories. Devices with compute compatibility less than 2.0, cannot read nor write to host memory directly. Data has to be explicitly transferred between CPU and GPU using runtime-libraries.

To maximise the utilisation of the available computing resources on the GPU, the minimum of two blocks must be equivalent to twice the number of SMs in the device. This is because running a single block per SM forces the SM to an idle state during synchronisation and global memory access. Therefore, scheduling two or more active blocks on each SM allows to switch between the blocks that are waiting and ready to be executed. Hardware hides the latency by switching the warps of the active blocks. Effectiveness of this switching depends on the occupancy of the SM. Hiding latency is better with higher occupancy. The state of the registers are not destroyed during the warp switching, thus this can be performed with no additional cost.

The threads scheduler selects a warp that is ready to be executed and issues the instruction to the active threads of the warps. Full potential is realised when a warp executes a single instruction concurrently on all of its threads. However, if threads within the warp diverge via a data dependent condition, then the entire warp is executed serially by predication. Branch divergence only take place within a warp so different warps execute independently regardless of whether they are executing common or disjointed code path [106]. Serialisation of the warp, carried out by thread predication, increase the number of instruction issues, thus reducing the performance. Every instruction in the serialised warp is fixed with a predicator, which determines and executes an instruction if necessary (i.e. threads are selected in order) [106, 111].

2.4 Conclusion

In conclusion, this chapter outlined an overview of OCT systems, fs laser inscription and GPU (CUDA) based data processing. A brief theoretical description of

CHAPTER 2. BACKGROUND

the OCT systems and fs lasers inscription were given. In addition to this, description of CUDA software and a possible hardware configuration were also presented.

Chapter 3

OCT phantom requirements

A femtosecond inscription method was developed for the fabrication of OCT calibration and characterisation phantoms. The femtosecond laser inscription process enables the production of complex three-dimensional structures, with sub-micrometer accuracy, within the bulk of a transparent medium. The motivations leading to the development of this technique are the high reproducibility and deterministic nature of material modifications achieved using fs lasers. Additionally, the properties of the modified structures such as dimension, position and refractive index can be verified with other methods such as an optical microscope. This leads to the rapid production and characterisation of the phantoms with high accuracy.

As discussed in §2.2, sub-micrometer structures can be fabricated within the bulk of a transparent material by focusing the fs beam inside the material and by translating the sample relative to the beam waist. Due to the non-linear nature of the laser matter interaction process, induced structures are confined to a small region around the beam waist where laser intensity is larger than the threshold intensity required for material modifications.

3.1 Resolution

The resolution of the system is one of many parameters that determines the performance of an OCT system [112]. The axial R_a and lateral R_t resolutions of an OCT system are decoupled and governed by physically different systems. Axial resolution is dominated by source coherence length, whilst the lateral resolution and the depth of focus are determined by the diffraction limit of the objective [31, 113, 114]. Therefore OCT can achieve high axial resolution independent of the lateral resolution (i.e. focusing optics). The lateral or transverse resolution, and the depth of field (DOF) over which the signal can be obtained, are related to the NA of the objective. A smaller focal spot and a higher transverse resolution is achievable with a higher-NA objective, but this decreases the DOF [115].

3.1.1 Rayleigh Criterion

Generally the resolution of an optical imaging system is defined in terms of the Rayleigh criterion, which states that two points are just resolvable when the angular separation equals the angular radius of the Airy disk $I(r)$ (i.e. peak intensity of one Airy disk overlays the first minimum of the second (shown in Figure 3.1)). The Airy disk is expressed by the following equation:

$$I(r) = \left(\frac{J_1(r)}{r} \right)^2,$$

where $J_1(r)$ is the Bessel function of the first kind and $r = \sqrt{(x + y)}$ is the orthogonal lateral dimension perpendicular to the optical axis z .

A Gaussian function satisfies the Rayleigh criterion when the superimposed fields yields a contrast of $\Delta I = 1 - 2e^{-1}$, hence the Rayleigh criterion determined OCT resolution is equal to the e^{-1} field radius of the focal spot size [1].

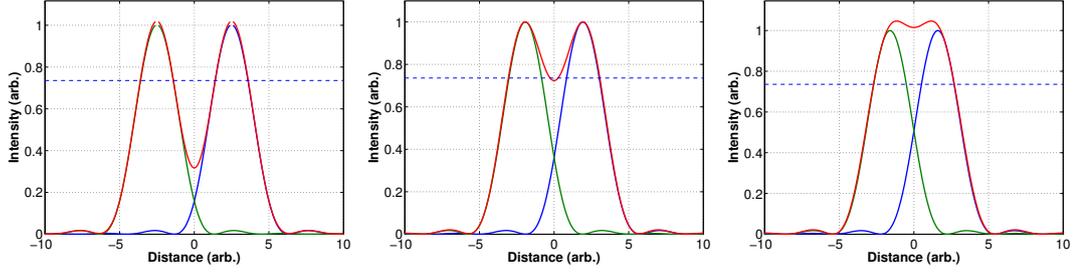


Figure 3.1: Airy disk separation: resolved (left), Rayleigh criterion (middle) and unresolved (right).

3.1.2 Lateral Resolution

The lateral resolution of the OCT is determined by the diffraction limit of the focusing optics as in a microscope. The diffraction limited spot size is inversely proportional to the NA or the focusing angle of the beam. The lateral resolution is then defined as [116]:

$$\Delta x = \frac{4\lambda f}{\pi d}, \quad (3.1.1)$$

with diameter of the spot size of the beam on the objective d and objective of focal length f . High NA objective can be used to achieve high axial resolution. The NA also determines the depth of focus (confocal parameter) $b = 2z_R$ where z_R is the Rayleigh range defined as [116]:

$$z_R = \frac{\pi \Delta x^2}{\lambda}. \quad (3.1.2)$$

Thus, the better the transverse resolution, the shorter the depth of field.

3.1.3 Axial Resolution

Generally in the OCT related literature, the axial resolution of the OCT system is defined as half the coherence length of the source (i.e. $l_c/2$). The coherence length is defined by the spatial auto-correlation measured by the detector [112]. The Wiener-Khinchin theorem states that the inverse Fourier transform of a power spectral density $S(\omega)$ is the auto-correlation function $\Gamma(\tau)$ [113, 117].

CHAPTER 3. OCT PHANTOM REQUIREMENTS

For an assumed Gaussian beam the power spectral density (PSD) is expressed as [113]:

$$S(\lambda) = \frac{2\sqrt{\ln(2)}\lambda_0^2}{\sqrt{\pi c\Delta\lambda}} \exp \left\{ - \left[2\sqrt{\ln(2)} \frac{(\frac{1}{\lambda} - \frac{1}{\lambda_0})}{\left(\frac{\Delta\lambda}{\lambda_0^2}\right)} \right]^2 \right\}, \quad (3.1.3)$$

and the autocorrelation function is given by:

$$\Gamma(\tau) = \exp \left[- \left(\frac{\pi c\Delta\lambda\tau}{\lambda_0^2 2\sqrt{\ln(2)}} \right)^2 \right] \exp \left[-i \left(\frac{2\pi c\tau}{\lambda_0} \right) \right], \quad (3.1.4)$$

where λ_0 is the centre wavelength and $\Delta\lambda$ is the spectral width of the source. From this the coherence length can be expressed as [23, 118, 119]:

$$l_c = \frac{4\ln(2)}{\pi} \frac{\lambda_0^2}{\Delta\lambda}. \quad (3.1.5)$$

Figure 3.2 illustrates the variation of axial resolution as a function of the full width at half maximum (FWHM) spectral width for three different central wavelengths (800 nm, 1300 nm and 1500 nm) predominantly used in OCT systems.

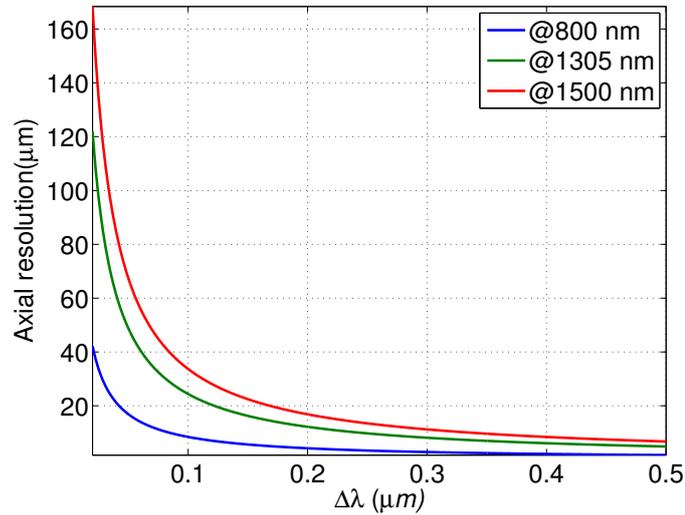


Figure 3.2: Simulated axial resolution (coherence length) as function of varying FWHM spectral width for three different central wavelengths (800, 1300 and 1500 nm).

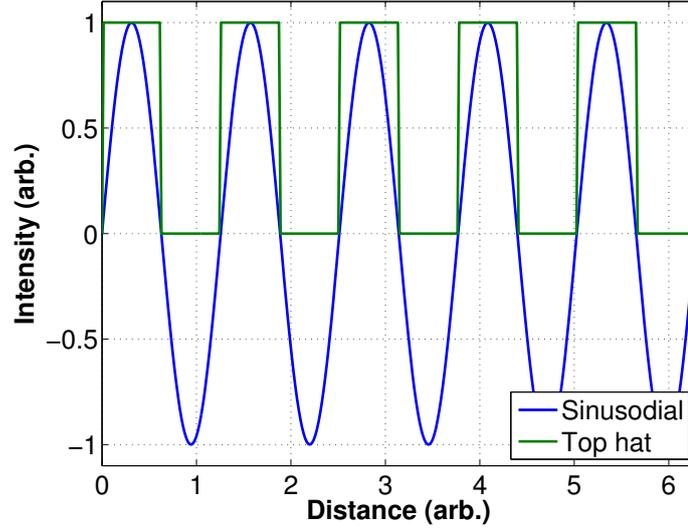


Figure 3.3: The simulated results of sinusoidal (MTF) and top-hat (contrast transfer function (CTF)) response functions.

3.1.4 Modulation transfer function

The modulation transfer function (MTF) determines the characteristic frequency response function of a system that is derived by taking the Fourier transform (FT) of the line spread function (LSF) or PSF. A sinusoidal object can be given as:

$$h_s(x, f) = \exp(i2\pi fx), \quad (3.1.6)$$

where x is wavelength/peak distance, f is the spatial frequency of the target in lp/mm . A fs inscribed structure can be approximated (in 2D) as a top hat function with a reflectivity pattern written as:

$$h_b(x) = \begin{cases} 0 & \sin(2\pi fx) \geq 0 \\ 1 & \sin(2\pi fx) < 0 \end{cases} \quad (3.1.7)$$

Figure 3.3 shows the simulated (2D) reflectivity profile of a sinusoidal (Eqn (3.1.6)) and top hat profiles (Eqn (3.1.7)).

The optical signal delivered by the detector (i.e. linescane camera or photodiode) of the OCT is the convolution of the object (input) function (h) with the

system response (PSF, g) given as:

$$I = h(x, f) \otimes g(x), \quad (3.1.8)$$

where g is the Gaussian function, $g(x) = \exp(-2i\pi x)$.

The normalised modulation M is defined as:

$$M = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}, \quad (3.1.9)$$

where I_{min} and I_{max} are the minimum and maximum measured intensities. The MTF is given as the ratio of the response to the input image to the response to the measured image:

$$MTF(f) = \frac{M_{img}}{M_{obj}} = \mathcal{F}\{g(x)\}. \quad (3.1.10)$$

3.2 Method

The PSF is one of the fundamental parameters that defines the three-dimensional resolution of an OCT systems. The final image $I(x, y, z)$ from any imaging system is the convolution of the systems PSF $g(x, y, z)$ and the object function $h(x, y, z)$ [120]:

$$\begin{aligned} I(x, y, z) &= g(x, y, z) \otimes h(x, y, z) \\ &= \int \int \int g(x, y, z) h(x - \chi, y - \psi, z - \zeta) d\chi d\psi d\zeta. \end{aligned} \quad (3.2.1)$$

Therefore, the PSF of an imaging system can be directly measured by imaging an object that can be approximated to a delta function i.e. something smaller than the resolution of the system. Conventional PSF measurements, such as imaging a mirror, provides a delta function in the axial direction [31].

When the cross section of the structures imaged is smaller than the resolution of the OCT system, the observed results can be approximated by the two-dimensional PSF of the OCT system. Hence direct measurement of the PSF is

CHAPTER 3. OCT PHANTOM REQUIREMENTS

possible. Also, the MTF or CTF of the system can be calculated by measuring the LSF in the axial and lateral direction and then taking their Fourier transformation. Additionally, customised phantoms can be produced to measure the spatial variation of the resolution (i.e. variation of resolution with axial and lateral positions), sensitivity, signal drop off as well as to verify the linearity of the scan. Another advantage of using PSF phantom is that the spatial variation of the PSF can be measured.

Writing linear features and taking B-scans normal to them results in points being seen in the B-scan image. Axial and lateral resolution and MTF can therefore be estimated by measuring PSF or LSF.

Imaging of three-dimensional objects/structures much smaller than system's resolution allows a direct measurement of the three-dimensional PSF [31, 114]. This technique has been used in other imaging systems and has been adapted for OCT elsewhere using a variety of approaches. Previously, particles embedded in clear resin [31] and suspended in aqueous solution [114] have been used successfully to measure the PSF of both time and Fourier domain OCT systems.

However, using particles embedded in a matrix has some practical constraints that affect the usability. Ideally, having a knowledge of all of the particle's properties such as size, reflectivity and location is desired. However, identifying the location (coordinate), size and reflectivity¹ of individual particles relative to a reference point is a time consuming task [114]. Therefore, statistical methods must be employed to measure the PSF. For an accurate measurement of PSF by statistical analysis, a homogeneous distributions of particles within the matrix is desired. However, inevitably, particles are not uniformly distributed within the matrix and potential aggregation of the particles will lead to a systematic spread in the results obtained. To minimize this issue, larger particles could be imaged and then their

¹Distribution in the size and reflectivity of the particles are comparable to the fs inscribed structures. The main advantage of a fs inscribed structure over the particles is a priori knowledge of the positions. This allows fabrication of identical phantoms, the particle phantom methods, there is no control over the positions of the particles in the matrix so each phantom will look different. Additionally, fs inscribed structures will also enable the creation of reference points, so the comparison of results between measurements becomes a much easier task.

known shape can be deconvolved to obtain the PSF values [31].

Due to the limitations of particle embedded phantoms, Tomlin et al. have demonstrated the applicability of using a fs micro-inscribed phantom in fused silica as PSF calibration sample [1]. Tomlins et al. demonstrated the feasibility of the technique for the production of OCT phantoms, but the structures exhibited prolonged elongation in the lateral direction and small variations along the structure resulting in variations in the measured intensity between B-scans [1]. Here we investigate the use of fs inscription to create such phantoms in PMMA and demonstrate the production of near cylindrical lines, which can be used to determine the PSF of the OCT.

3.3 Point spread function

To measure the three dimensional resolution, infinitesimally small (delta function like object response i.e. $\Pi(x) = \delta(x - x_0)$) three dimensional objects, such as nano-particles, can be used. In practice, fs laser inscribed structures can be considered to have sub-micrometer dimensions. Understanding the effect of finite structure size of the phantom on the PSF width allows one to determine the maximum size of the phantom structures that can be used to measure the PSF. In addition, structures used for determining the PSF must create a large amount of back-scattering so that they can be detected by the OCT systems. This in effect sets the lowest limit on the size and reflectivity of the structures. In order to create maximum visibility by any system, it is better to create structures near the upper size limit.

To estimate the upper limit of the structure size that has negligible contribution to the PSF broadening effect, let us consider the phantom as a 2D object. The inscribed structures can then be modelled in two dimensions using a top hat function [31]:

$$\Pi(x) = \Phi(x + \Delta/2) - \Phi(x - \Delta/2), \quad (3.3.1)$$

CHAPTER 3. OCT PHANTOM REQUIREMENTS

where Δ is the finite cross-sectional size of the structure (intimater) and $\Phi(x)$ is the unit step function, defined as:

$$\Phi(x) = \begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases} \quad (3.3.2)$$

The convolution of a particle of finite size with the PSF in equation (3.3.1) of the system can be analytically expressed as [31]:

$$g(x, y, z) \otimes \Pi(x) = \sqrt{\frac{\text{erf}\left[\frac{1}{2}(\Delta - 2t)\sqrt{2/W_t^2}\right] + \text{erf}\left[\frac{1}{2}(\Delta + 2t)\sqrt{2/W_t^2}\right]}{2\text{erf}\left[\frac{1}{2}\Delta\sqrt{2/W_t^2}\right]}}, \quad (3.3.3)$$

where $\text{erf}(x)$ is the error function and \otimes is the convolution operator.

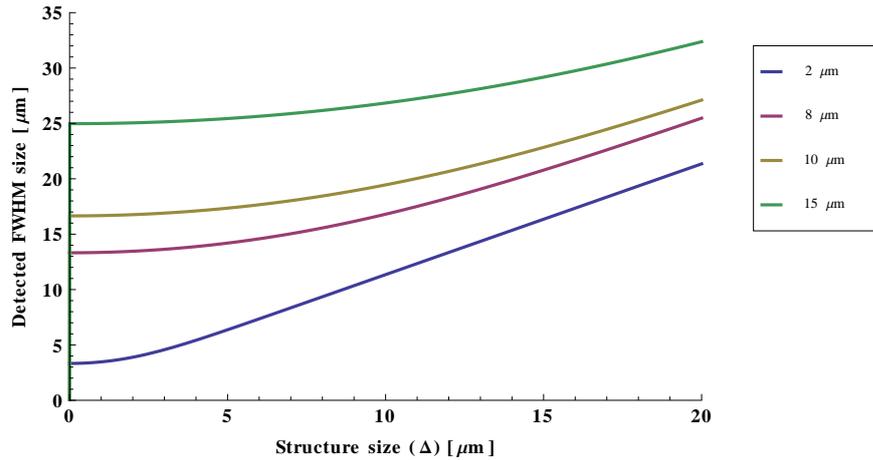


Figure 3.4: Simulation results for detected particle size, obtained with equation (3.3.3), by OCT with various resolutions, for particles with varying size.

Figure 3.4 shows the variation in the simulated detected structure size for convolving structures of varying size with PSF in of FWHM widths of 2, 8, 10 and 15 μm . Ultra high resolution OCT system have been reported to achieve better than 2 μm resolution. To experimentally measure the resolution of these system, structures therefore need to be much smaller than 1 μm . On the other hand conventional systems (i.e. all the system used in this study) have much lower

resolution. For a typical OCT system with a $8 \mu m$ lateral resolution, structures of less than $3 \mu m$ in (lateral) size would be acceptable.

3.4 Focusing conditions

The fs laser inscription involves tightly focusing the laser beam within the bulk of the media at the place where the structures to be created are required. The fs laser inscription is a non-linear process, and only at the focus there is enough intensity to produce structural modification. Inscribed structures are therefore confined to the focal region, and the dimension of the region is related to the focusing conditions. In this section we investigate the requirement of the focusing condition necessary to achieve the optimal dimension required for OCT phantoms.

The size of a diffraction limited focused spot is given by:

$$\omega_0 = \frac{0.61\lambda}{NA}, \quad (3.4.1)$$

where ω_0 is the beam waist after focusing a beam with centre wavelength λ by an objective with the NA. The Rayleigh length z_R relates to the dimension of the structure in the axial direction of the beam, it is expressed as:

$$z_R = \frac{\pi\omega_0^2 n}{\lambda}. \quad (3.4.2)$$

The size of the modified region depends on the NA and the wavelength of the laser (λ). In the ideal case, the axial and lateral dimension of the inscribed structure are proportional to Rayleigh length (confocal parameter) and the beam waist respectively.

Neglecting the effects of the self-focusing, the lateral and axial dimensions of the induced structures by fs laser near the surface of the sample can be approximated by radius of the focal spot ω_0 and the confocal parameter b [45]. These

can be expressed as:

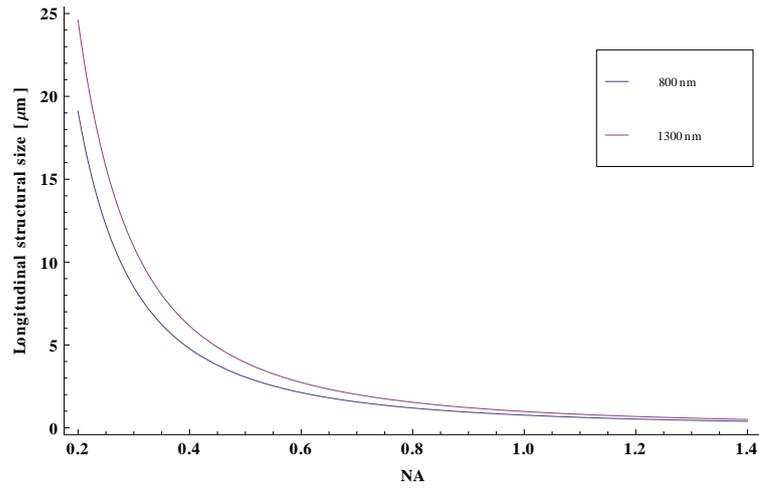
$$\omega_0 = \frac{M^2 f \lambda}{\pi \omega_s}, \quad (3.4.3)$$

$$b = 2 \frac{M^2 f^2 \lambda}{\pi \omega_s}, \quad (3.4.4)$$

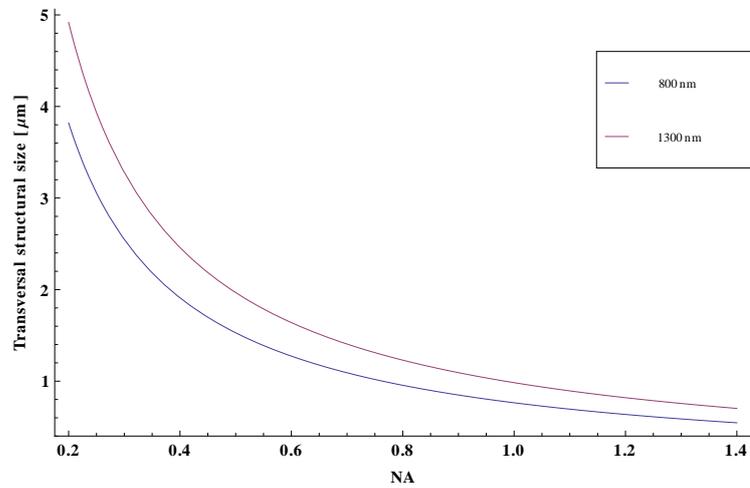
where w_s is the radius of the beam at the objective, λ is the wavelength, f is the focal length and M^2 is a dimensionless parameter which quantify the quality (divergence) of the real beam relative to a diffraction-limited Gaussian beam. For an ideal Gaussian beam TEM_{00} , this ratio becomes unity. The NA of an objective can be approximated to $NA = \frac{w_s}{f}$. Figure 3.5 shows the variation of the dimension of the structures inscribed using 800 nm and 1030 nm wavelength lasers for varying NA.

In practice, a number of factors such as self-focusing halt the optimum focusing inside the bulk of the material. Therefore, this simple approximation may become invalid when the laser is focused inside the bulk of the transparent material. Processes such as self-focusing and spherical aberration must be taken into account to predict the actual size. Nevertheless, this simple approximation provides an indication of the NA required to produce the optical structures (dimension) for OCT calibration/characterisation phantoms. Figure 3.5 indicates that to measure axial and lateral dimension of an OCT with a resolution of 2 μm , structures need to be smaller than 1 μm and for an OCT with 8 μm resolution, structures smaller than 3 μm are required.

CHAPTER 3. OCT PHANTOM REQUIREMENTS



(a)



(b)

Figure 3.5: Longitudinal and transverse size of the beam as function of NA for central wavelengths 800 nm and 1026 nm.

3.5 Sensitivity

The sensitivity of the system is defined as the reciprocal of the minimum reflectivity of the sample that the system can detect with signal to noise ratio (SNR) of one, $\Sigma = R_{S,min}^{-1}$ [27, 121]. An artefact with varying reflectivity can be used to determine the sensitivity of the system. The sensitivity is defined as the smallest intensity signal that a system can detect which is just above its threshold for de-

tection. The sensitivity of an OCT system is determined by several parameters, including the sensitivity of the detector, optical power of the light source, imaging depth, the resolution limit of the spectrometer and aliasing (or reduced fringe visibility) [27, 28]. SD-OCT systems suffer from sensitivity fall-off along axial direction. Also optical aberrations can hinder the sensitivity.

Conventionally, the sensitivity of an OCT system is determined by measuring the signal-to-noise ratio of a known reflector. A mirror or a piece of optical glass with a known refractive index is commonly used in combination with a tilted reflective-type neutral density filter. To obtain accurate results, the reflectivity and transmission of the phantoms must be accurately measured. The sensitivity can be expressed as:

$$SNR = 20\log_{10} \left(\frac{P}{N} \right) + 10\log_{10} (RT^2), \quad (3.5.1)$$

where R is the mirror reflectance, T is the transmission of the optical filter and the P and N are the signal and noise levels respectively measured from the linear processed OCT data.

The amount of back scattered light is proportional to the refractive index change and the scattering area of the structure. The structures in the sensitivity phantom were inscribed with varying size and reflectivity; therefore, these can be used to identify the sensitivity limit of the system. The sensitivity limit of OCT corresponds to the smallest/faintest structure that is resolvable by the OCT.

Structures with varying reflectivity can be produced inside a transparent material by fs laser inscription by varying the speed or power of the laser. From Fresnel equations, the reflectivity \mathcal{R} of a normal incident light at the boundary with different refractive indices n_1, n_2 is given as:

$$\mathcal{R} = \left| \frac{n_2 - n_1}{n_2 + n_1} \right|^2. \quad (3.5.2)$$

This indicates a square relationship between the increase in the reflectivity and

the refractive index difference between the mediums. This implies that microstructures with varying refractive indices can be used for measuring the sensitivity of an OCT system.

3.6 Distortion phantom

The geometrical distortion in an OCT image can arise from many sources such as optical aberration and scanning methods/geometries. Theoretical models have been developed over the years to compensate for these distortions [122, 123]. To date, only numerical based methods have been reported for quantification and correction of distortions in OCT systems.

Phantoms with ‘control points’ in three dimensions have previously been used for measuring the distortion in magnetic resonance imaging (MRI) systems [124, 125]. The principles used in MRI distortion phantoms can be applied to OCT distortion phantoms. Development of a suitable phantom for the quantification of three-dimensional distortion is essential.

The concept of a distortion phantom is quite simple; it contains a set of features called ‘control points’ that can easily be identified in the image data. Distortion in the image domain is measured by establishing a correspondence between the identified control points in the image and their known position. Distortion measurement and correction using phantoms can be summarized as follows: identify the control points (i.e. intersections) in the image space and then determine the distortion tensor. The tensor is found by calculating and comparing the deviation of the control points in the distorted image space to their corresponding points in the world (phantom) space. The distortion tensor is used to generate a mapping function that maps the distorted space to the world space. This can be represented by linear terms (scaling and shear) and high order non-linear terms [124]:

$$\vec{u} = F(x) = t + \sum_{i=1}^4 a_i x^i + O(x^5), \quad (3.6.1)$$

CHAPTER 3. OCT PHANTOM REQUIREMENTS

where $\vec{u} = (u, v, w)^T$ and $x = (x, y, z)^T$ are the coordinates points in undistorted world space (phantom domain) and distorted image space respectively. $\vec{t} = (t_x, t_y, t_z)^T$ is the translation vector, where a_i (where $i = [1, 4]$) are the transformation matrices containing i^{th} order transformation coefficients. The mapping function is used to correct the distortion in the image space.

For a three dimensional imaging system, 2D phantoms (control points in x, y plane) can only provide information in 2D, since the third dimension (z) is unknown/undefined. To measure the 3D distortion of an OCT system, three-dimensional distortion phantoms (i.e. control points in 3D space) are required. The relationship between these control points in a phantom and the distorted image can be described as follows:

$$\begin{pmatrix} x \leftrightarrow x' \\ y \leftrightarrow y' \\ z \leftrightarrow z' \end{pmatrix} \quad (3.6.2)$$

The geometric distortion within the effective volume phantom using positional deviation, obtained using interpolation methods from control points in phantom space to those in distorted image space, can be written as [124]:

$$dr_{ijk} = r'_{ijk} - r_{ijk} \quad (3.6.3)$$

where $dr(r)$ is the distortion, r' and r are the coordinates of the intersections in image and real/phantom space respectively.

The control points required for a distortion phantom can be created in one of two ways, either by placing individual microstructures at discrete locations or line intersections. We opted for the latter (i.e. line intersections) due to the ease of fabrication but are currently investigating individual points.

3.7 Oversampling and Smoothing

The spatial resolution of the OCT data can be improved by oversampling the data. Aligning the inscribed structures non-orthogonal OCT beam axis, by rotating the sample, gives an offset between different B-scans. As described by Wang et al. [126], the position of the lines vary in a predictable manner from one B-scan to the next thus incorporating the data obtained in multiple B-scans provides an oversampled representation of the detected point at the Nyquist limit.

Wang et. al. used a method based on the Fourier shift theorem to determine the offset of each line. According to the Fourier shift theorem, a delay in spatial or time domain is equivalent to a linear phase term in frequency domain [126]:

$$X(t - \tau) \implies \exp(-i\omega\tau)X(\omega), \quad (3.7.1)$$

where $X(\omega)$ is the FT of the $x(t)$.

Woolliams and Tomlins have shown an alternative approach that estimates the centre of the points that will be used in this study. This is more straightforward to implement and seems to give comparable results [127]. The positions of the peak of each point was computed by fitting a two-dimensional Gaussian function to a region around the cross-section in B-scan and then co-registering the points to a common coordinate and smoothed using a Gaussian smoothing Kernel. Alternatively, co-registered data were interpolated and then spatially averaged.

3.7.1 One dimension

In an en-face or side-view image, each row is essentially a line profile (lateral) that can be represented as a set of PSFs. We can represent the measured intensity values of a line (of discrete points) as a set $\{PSF_p\}$ where $p = [1, P]$. Then we can consider that the *en-face* image is composed of multiple PSFs, which can be represented as multiple subsets $\{PSF_p\}_q$, where $q = [1, Q]$ represents the q^{th} row and Q is the maximum individual PSF measurement (i.e number of lines/B-

scans).

The PSFs in *en-face* (and side-view) image are co-registered to a common coordinate system $\{x'_p\}_q$ as follows:

$$\{x'_p\}_q = \{x_p\}_q + \{x_0\}_q. \quad (3.7.2)$$

Co-registered data were binned by one forth the pixel size and averaged to give a four time oversampled data.

3.7.2 Two dimensional data

The cross-section of the points in the B-scan corresponds to a set of intensities values at discrete coordinates $\{r_p, z_p\}$, where $p \in [1, P]$ represents the measured p th measurement and P is the total number of the discrete intensity of measurements of $PSF(r, z)$ [128].

The centre $\{x_0, y_0\}$ of the $\{PSF_p\}_q$ was estimated by fitting a two dimensional Gaussian function of the form given in equation (A.0.1) using the least square method described in appendix A. The PSF of the peaks was off-set by $\{x, y\}_q$, and the resulting new coordinate is given by:

$$\{x', y'\}_q = \{x, y\}_q + \{x_0, y_0\}_q, \quad (3.7.3)$$

where $\{x', y'\} \subset \{x, y\}$. A common coordinate $\{x, y\}$ was formed by aggregating all the PSF_q . Alternatively, the centroid of the peak can be used for co-registering the PSFs. Computing the centroid of each frame from an extracted sub-volume, is computationally less extensive than fitting a Gaussian function.

A two dimensional Gaussian kernel smoother (moving weighted average) given in equation (3.7.4) was used to smooth and average the data [128]:

$$PSF_{smooth}(r_K, z_K) = \frac{\sum_{j=0}^{J-1} K(r_j - r_K, z_j - z_K) PSF_j}{\sum_{j=0}^{J-1} K(r_j - r_K, z_j - z_K)}. \quad (3.7.4)$$

3.8 Deconvolution

The image acquired from any imaging system is a convolution of the system response and the object function [120, 129]. Previously, mathematically based deblurring [31, 114, 129, 130], experimentally based point spread function measurements [31, 114, 120, 130] and spatial deconvolution were demonstrated [120].

The convolution of an image with the system response, given in equation 3.2.1, can be expressed in the Fourier domain as:

$$F(u, v) = G(u, v)H(u, v), \quad (3.8.1)$$

where F,G and H represent the measured image, PSF and the original image of the OCT respectively, all in the Fourier domain.

Previously, several groups have reported the achievement of a considerable resolution improvement by deconvolution. It has previously been shown that enhancement to the resolution improvement can be realised by means of post processing. Deconvolution was used by Kulkarni et al. to overcome the longitudinal resolution limit [131]. Schmitt et al., Wang et al. and Liu et al. have performed two dimensional deconvolution by applying both axial and lateral PSF [129, 132, 133]. Wang et al. proposed a combined Wiener filtering with homomorphic filtering for spectral and spatial deconvolution respectively [133]. They all considered the PSF as a spatially invariant entity and used a single PSF for an entire image estimated from source coherence function and Gaussian beam optics.

The PSF of an OCT system can be considered as spatially invariant when $g(x, \zeta) = g(x - \zeta)$. However, the resolution of an OCT system varies with axial and lateral position due to the variation in the objective and detection system employed in SD-OCT systems. Agrawal et al. [114] and Woolliam et al. [120] have measured the spatially variant PSF of an OCT using particle phantoms. The B-scan can be split into $(m \times n)$ sections and the PSF within each section can be assumed to be spatially invariant. Each section can be deconvolved using the

corresponding PSF. Woolliams et al. have created a spatially variant point spread function map (PSF-map) by segmenting the image into small regions where the resolution was considered as spatially invariant [120]. The measured PSF was used for deconvolution corresponding region of an image from another set of data [120].

Deconvolution is a computationally expensive process, which limits the realisation of its potential for real time post-processing to improve the resolutions. Previously GPU based data processing was shown to achieve incredible speed that allows real time processing and rendering of the OCT data. At such high processing speeds, it is feasible to compromise some of the processing speed to incorporate more advanced signal and image processing techniques for enhancing OCT data. Potential of real-time deconvolution can be realised by GPU accelerated techniques.

3.8.1 Theory

The expected interference signal $h(k)$ of an OCT can be expressed as [133]:

$$h(k) = s(k) |1 + A(k)|^2, \quad (3.8.2)$$

where $s(k)$ is the source spectrum and $A(k)$ is the Fourier Transform of $a(z)$,

$$A(k) = \int_0^\infty a(z) e^{i2kz} dz, \quad (3.8.3)$$

The observed signal $f(k)$ is degraded by the additive $\eta_a(k)$ and multiplicative $\eta_m(k)$ noise. In addition it is also convoluted by the system response (PSF) $g(k)$ which arises from the characteristics of the light source and the imaging optics employed. Hence the detected signal in SD-OCT can be expressed as [133]:

$$f(k) = g(k) \otimes h(k) \eta_m(k) + \hat{\eta}_a(k), \quad (3.8.4)$$

CHAPTER 3. OCT PHANTOM REQUIREMENTS

where $\hat{\eta}_a = g(k) \otimes \eta_a(k)$. Depth profiles of the sample is obtained by taking FT of $f(k)$ with respect to k

$$F(z) = \mathcal{F}\{f(k)\} = G(z)H(z) \otimes N_m(z) + \hat{N}_a(z), \quad (3.8.5)$$

$$\hat{N}_a(z) = G(z)N_a, \quad (3.8.6)$$

where $H(z)$ contains the depth profile information, $G(z)$ is the optical transfer function of the system, N_m and N_a are the Fourier transform of η_m and η_a respectively [133].

The presence of noise in the measured data makes difficult to separate the true image from the noise by direct calculation, hence indirect or iterative processes are required to minimise the noise. Deconvolution algorithms are iterative procedures that try to account for the statistical properties of the noise. Several algorithms have been developed over the years of which the LR and WF are the most widely used algorithms.

Wiener filter (WF)

The Wiener algorithm is based on least square approximation method, which can be written in Fourier domain as [129],

$$H(u, v) = F(u, v) \left[\frac{G(u, v)}{G(u, v)G^*(u, v) + \eta(u, v)} \right], \quad (3.8.7)$$

where $H(u, v)$, $F(u, v)$ and $G(u, v)$ represent the spectrum of the original image (ideal image), spectrum of the measured (blurred) image and the 2D optical transfer function (OTF) of the OCT system respectively. $\eta(u, v)$ was the approximate spectrum density of noise $S_\eta(u, v)$ to the signal $S_f(u, v)$ ratio (this is $1/\text{SNR}$):

$$\eta(u, v) \approx \frac{S_\eta(u, v)}{S_f(u, v)}. \quad (3.8.8)$$

Lucy-Richardson (LR)

The LR algorithm given in equation (3.8.9) uses Poisson statistics to model the noise and provides the maximum likelihood estimate for the true image [129]

$$h_{k+1}(x, y) = h_k(x, y) \left[g(-x, -y) \otimes \frac{h(x, y)}{g(x, y) \otimes h_k(x, y)} \right], \quad (3.8.9)$$

where $h_{k+1}(x, y)$ is the estimate of the original image after k^{th} iteration and $f(x, y)$ and $g(x, y)$ are the degraded image and the PSF of the imaging system.

In order to satisfy the spatial variant of the resolution in OCT data, we use the experimentally measured PSF for deconvolution. Detailed description of measuring spatially variant PSF of an OCT system using the OCT-phantoms are given in section 7 and elsewhere [31, 114, 120]. The B-scans are divided into $(m \times n)$ sub-regions $\Delta_{x,y}^{(m,n)}$ where corresponding PSF $P_{x,y}^{(m,n)}$ will be used for deconvolution. The B-scan is split into $m \times n$ subsection $D \in R^{M \times N}$ and each sub image is $\Delta \in R^{I \times J}$ where m and n denote the row and column index of the sub-image and i and j denote the row and column index of the pixels within a sub image. In this respect each subsection can be deconvolved independently using the corresponding PSF. Each sub-image can be treated as a set and rewritten using the deconvolution algorithm for entire B-scan as follows:

$$\{F(u, v)\}_{(m,n)} = \{G(u, v)\}_{(m,n)} \left[\frac{\{H(u, v)\}_{(m,n)}}{\{H(u, v)\}_{(m,n)}\{H^*(u, v)\}_{(m,n)} + \{\eta(u, v)\}_{(m,n)}} \right], \quad (3.8.10)$$

$$\{f_{k+1}(x, y)\}_{(m,n)} = \{f_k(x, y)\}_{(m,n)} \left[\left\{ \{h(-x, -y)\}_{(m,n)} \right\} \otimes \frac{\{g(x, y)\}_{(m,n)}}{\{h(x, y)\}_{(m,n)} \otimes \{f_k(x, y)\}_{(m,n)}} \right], \quad (3.8.11)$$

where (n, m) represent the index of the sub-region. This allows us to apply the deconvolution to all subsections concurrently.

3.8.2 Tiled deconvolution

The deconvolution process is based on the assumption that the blur is approximately spatially invariant within the small regions of the image domain. There were three types of deconvolution methods that have been demonstrated for restoring the images using spatially variant PSFs.

- **Geometrical coordinate transformation** [134–138], where a spatially variant PSF is transformed into a spatially invariant, by coordinate transformation of 'distortion' or by other known symmetries. Restoration is applied to this transformed image, and a final transformation is used to obtain the final restored image.
- **Sectioning method** [135, 138, 139], where the PSF is assumed to be spatially invariant within the small region of the image domain. Image is partitioned into small tiles, and each tile is restored using its corresponding spatially varying PSF. Finally, individual restored tiles are then stitched together to form the final restored image. This could create artefacts near the boundary of the tiles. This is avoided by using *overlap-save method (OSM)* or *Overlap-Add method (OAM)*, and a descriptions of these methods are given below:
 - *OSM*, where the input data $D(x, y)$ (entire B-scan) is zero padded at the start to avoid wrap-around pollution. Zero padded input data is tiled, where each tile is overlapped slightly, after deconvolution overlapped regions are discarded.
 - *OAM*, where the input data $D(x, y)$ is split into non-overlapping tiles, which are then zero padded. Overlapping regions of the (zero padded) tiles are added together.
- **Interpolated PSF** [135, 138]- similar to the sectioning method, where the PSF is assumed to be spatially invariant within the small region. But rather

than deconvolving each tile separately with its corresponding PSF, the PSF of each tiles are stitched together using an interpolation method and then the interpolated PSF is used for deconvolving the full image.

3.8.3 Quantification of improvement by deconvolution

The quantities: local contrast (LC), contrast-to-noise-ratio (CNR), average SNR (SNR_{av}) and sharpness were computed to quantitatively estimate the image enhancements. The LC, CNR and SNR_{av} were defined as follows [133]:

$$LC = \frac{\mu_0}{\mu_b} \quad (3.8.12)$$

$$CNR = 20 \log_{10} \frac{\mu_0 - \mu_b}{\sigma_b} \quad (3.8.13)$$

$$SNR_{av} = 20 \log_{10} \frac{S_e}{N_e}, \quad (3.8.14)$$

where μ_0 and μ_b are the mean value of the object region and of the background respectively. These were estimated as follow [133]:

$$\mu_0 = \frac{1}{n_0} \sum d_{ij} \iff (d_{ij} \geq d_{th}) \quad (3.8.15)$$

$$\mu_b = \frac{1}{n_b} \sum d_{ij} \iff (d_{ij} < d_{th}), \quad (3.8.16)$$

where d_{ij} are the intensity values in the sub region extracted around the structure, and n_0 and n_b are the number of pixels, belonging to the object and background. The regions with intensity values are greater than a given threshold d_{th} , are taken as the object region and the rest is taken as background. The standard deviation of the background noise is given by σ_b . The mean energy of the objective S_e and mean energy of the background N_e are given by [133]:

$$S_e = \sqrt{\frac{1}{n_0} \sum d_{ij}^2} \iff (d_{ij} \geq d_{th}) \quad (3.8.17)$$

$$N_e = \sqrt{\frac{1}{n_b} \sum d_{ij}^2} \iff (d_{ij} < d_{th}). \quad (3.8.18)$$

CHAPTER 3. OCT PHANTOM REQUIREMENTS

The quantitative image sharpness was estimated using the equation (3.8.19) [129]

$$K = \frac{I_{peak}}{I_{side}}, \quad (3.8.19)$$

where I_{peak} and I_{side} are the intensity of the peak and of the average of the neighbouring values around the peaks, calculated using the equations (3.8.20) and (3.8.21). The sharpness of the edges in the images are proportional to K . Figure 3.6 illustrates the sharpness calculation where I_{peak} is at the centre $(0, 0)$ (black) and the I_{side} is indicated by dark grey colour surrounding the peak.

$$I_{peak} = P(x_0, y_0) \quad (3.8.20)$$

$$I_{side} = \sum_{\substack{i,j=-1, \\ i \neq 0 \\ j \neq 0}}^1 \frac{1}{8} I(x_0 + i, y_0 + j), \quad (3.8.21)$$

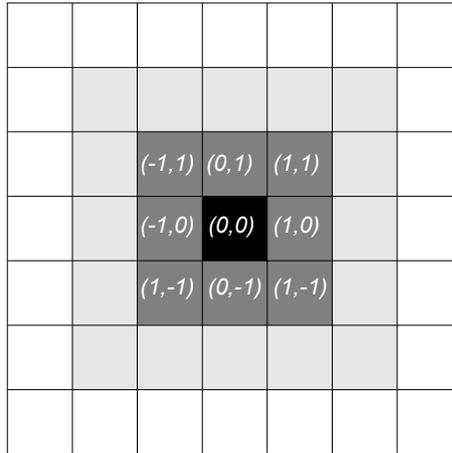


Figure 3.6: Example of calculating the sharpness. I_{peak} is at the centre $(0, 0)$ (black) and the I_{side} is indicated by dark grey colour surrounding the peak.

3.9 Conclusion

In conclusion, this chapter described the requirements for an ideal *OCT-phantom* to serve assessment of the system resolution, sensitivity, distortion and other pa-

CHAPTER 3. OCT PHANTOM REQUIREMENTS

rameters along with the description of these parameters. In addition to this, deconvolution methods were also discussed.

Chapter 4

Experimental set-up

4.1 Femtosecond laser

Three different fs laser systems were used for parameter space tests and fabrication of OCT calibration artefacts. In addition to the different operating conditions of these laser systems, the experimental set-ups also differ greatly from one another which influences the properties of the inscribed structures and the dimensions of the phantoms. For example, *Low-Rep 800 nm* system was designed primarily for the fabrication of fibre Bragg gratings. Thus, this is not ideal for mounting planer samples. The following sections describe these systems in details.

The following section describes the experimental set-up of the three different fs laser systems used in this study.

4.1.1 Lower repetition rate system: 800 nm

The lower repetition rate system is a Ti:Sapphire¹ chirped pulse amplification laser system consisting of a Tsunami oscillator and a Spitfire amplifier (both from Spectra Physics, Mountain View, CA). This laser system operates at a centre wavelength of 800 nm and produces 110 fs long pulse with a repetition rate of

¹Titanium doped sapphire

CHAPTER 4. EXPERIMENTAL SET-UP

1 kHz and a maximum pulse energy (average power) of 0.8 mJ (1 W). Here onward this system will be denoted as *Low-Rep 800 nm*.

The Tsunami mode locked Ti:Sapphire laser operates within a tuning range of 735 – 840 nm (central wavelength 790 nm), with an average power of 500 mW , a peak power greater than 75 kW , a pulse width less than 130 fs and a repetition rate of 82 MHz . The beam diameter and divergence are less than 2 mm and less than 0.6 $mrad$ respectively. This laser is pumped by a Millennia X (Spectra Physics, Mountain View, CA) diode pumped cw laser at 530 nm that provides $> 10 W$. The pulses produced by Tsunami are amplified by a Spitfire chirped pulse amplifier from nJ to mJ . An Evaluation (Spectra Physics, Mountain View, CA) Nd:YLP² frequency doubled (intra-cavity doubled, frequency dumped) laser that produces a Q-switched pulse with an average power of 6 W centred at 527 nm with a repetition rate of 1 kHz was used for pumping the Spitfire amplifier.

The pulse energy of this fs laser system can be adjusted using a half-wave plate and a Glan-prism. The half wave plate rotates the linearly polarised wave, and the Glan-prism only transmits the vertically polarised component of the beam. Therefore, the energy of the laser beam can be varied by rotating the half-wave plate without altering the polarisation of the beam. A rotator stage (M-036-DG, Physickinstruments) is used for rotating the polariser. The rotator can be controlled within the *G-code* by executing a dynamic-link library (DLL) wrapper written in C++³. The energy of the inscription beam was continuously measured using a thermal power-meter (FieldMaxxII-TO, Coherent) equipped with 1 mW – 3 mW thermal head. An auto-correlator (Model 80 - 50 Grenville, Swamp Optics) was used to measure the pulse duration. Quality of the pulses M^2 was measured using a M^2 – 200S meter (Laser Lines) and a spectrometer was used for measuring the laser spectrum.

²Neodymium-doped yttrium lithium fluoride

³Functions were written by myself.

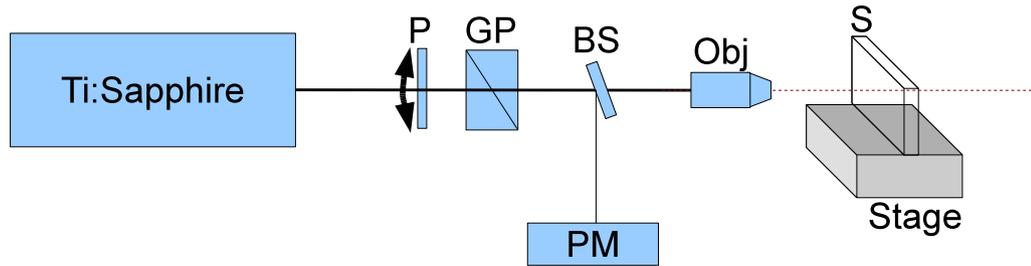


Figure 4.1: Experimental set-up of LR800 fs laser system. *P*: polariser; *GP*: Glen prism; *BS*: beam splitter; *Obj*: focusing objective; *S*: sample; *PM*: power meter.

The samples were mounted on a computer controlled five-dimensional (3 linear + 2 rotation) translation stage. The five dimensional translation stage is made by combining a two dimensional linear (Aerotech ABL1000) air bearing translational stage with three dimensional mechanical stage that has one linear and two rotational movements. The air-bearing translation stage has a resolution of 0.2 nm and repeatability of $\pm 50 \text{ nm}$ in both axis. These have travelling range of 150 mm and are able to reach a maximum speed of 300 mm/s with maximum acceleration of 10 mm/s^2 . The mechanical stage has 2 nm resolution linear translation and 0.027 to 0.057 mrad resolution in the rotational axis. The range of the linear and rotation axes are 4 mm and 20° respectively. All stages were controlled through a control unit (Aerotech A3200) connected to a personal computer (PC). The stage motions were controlled using a control unit that converts numerical commands, written in *G-code*, into analogue/mechanical commands and drive the stages.

4.1.2 Lower repetition rate system: 1026 nm

The second laser system used for this study is a Yb:YAG⁴ mode-locked fs laser (Amplitude Systemes s-Pulse HP) with a centre wavelength of 1026 nm with an average pulse duration below 500 fs . The laser has a tunable pulse repetition rate between $1 - 100 \text{ kHz}$, and can deliver pulses with maximum power of 4.1 mW (maximum pulse energy of 1 mJ). Figure 4.2 shows the experimental set-up of this system.

⁴Ytterbium (Yb) doped yttrium aluminium garnet (YAG)

CHAPTER 4. EXPERIMENTAL SET-UP

The samples were mounted on a goniometer with two axis of freedom to allow for tilt correction during the alignment of the samples. The goniometer, with the sample, was secured on a three axis linear translational stage (Aerotech ABL1000). This stage is same as that used in *Low-Rep 800 nm* system with two air-bearing translation stage and a mechanical Z translation. In addition, another mechanical linear translation stage was used to control the focus of the machining lens and an off-axis camera during the alignment. This was not used during the inscription processes. Properties of the stages are summarised in the table 4.1.

Table 4.1: *Properties of the translational stage*

Axis name	x	y	zz	z
Travel (<i>mm</i>)	150	50	4	100
Resolution (<i>nm</i>)	1	1	2	1000
Repeatability (<i>nm</i>)	50	50	50	1000
Accuracy (<i>nm</i>)	500	200	200	3000
Max speed (<i>mm/s</i>)	300	300	50	50

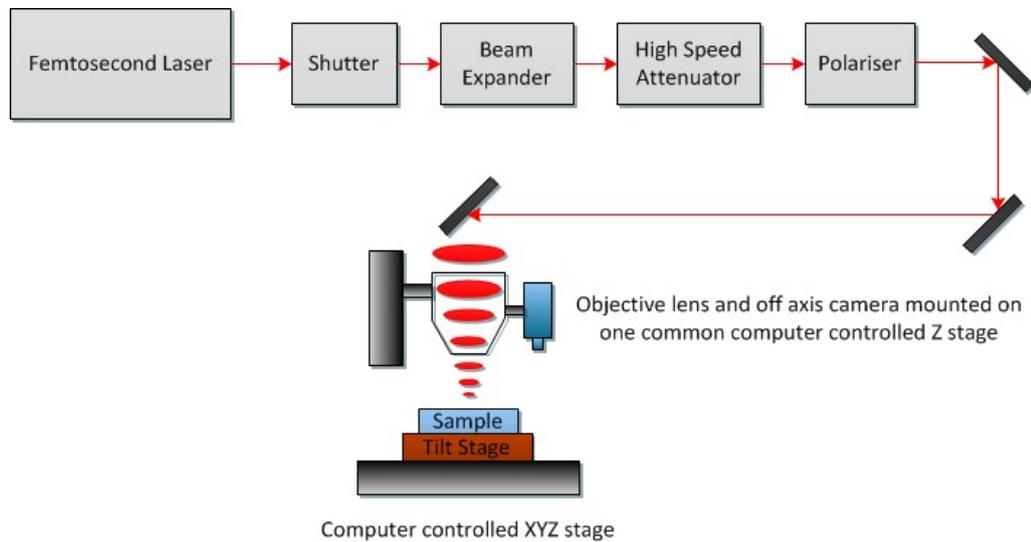


Figure 4.2: *Experimental set-up of the Low-Rep 1026 nm system.*

A pair of polarisers was used for controlling the output pulse energy as described in system 1. The relationship between the relative position of the polarisers and the maximum energy output is non-linear. The polariser was rotated

CHAPTER 4. EXPERIMENTAL SET-UP

relative to each other (i.e. 100% means all laser power transmitted through the polariser and 0% mean laser is completely blocked) and the average power was measured, after focusing the beam using a $100\times$ objective, using a power meter (Coherent FieldMax II mW) as function of the relative position of the polarisers. Figure 4.3 shows the measured *Low-Rep 1026 nm* laser power at the objective as function of relative polariser position. A simple quadratic function was fitted to the measured data which allows to estimate absolute energy from the percentage of the polariser rotation as follows:

$$y = -2.275 \times 10^{-3}x^2 + 1.759x + 4.016, \quad (4.1.1)$$

where y is the expected optical output of laser corresponding to the relative position of the polarisers x .

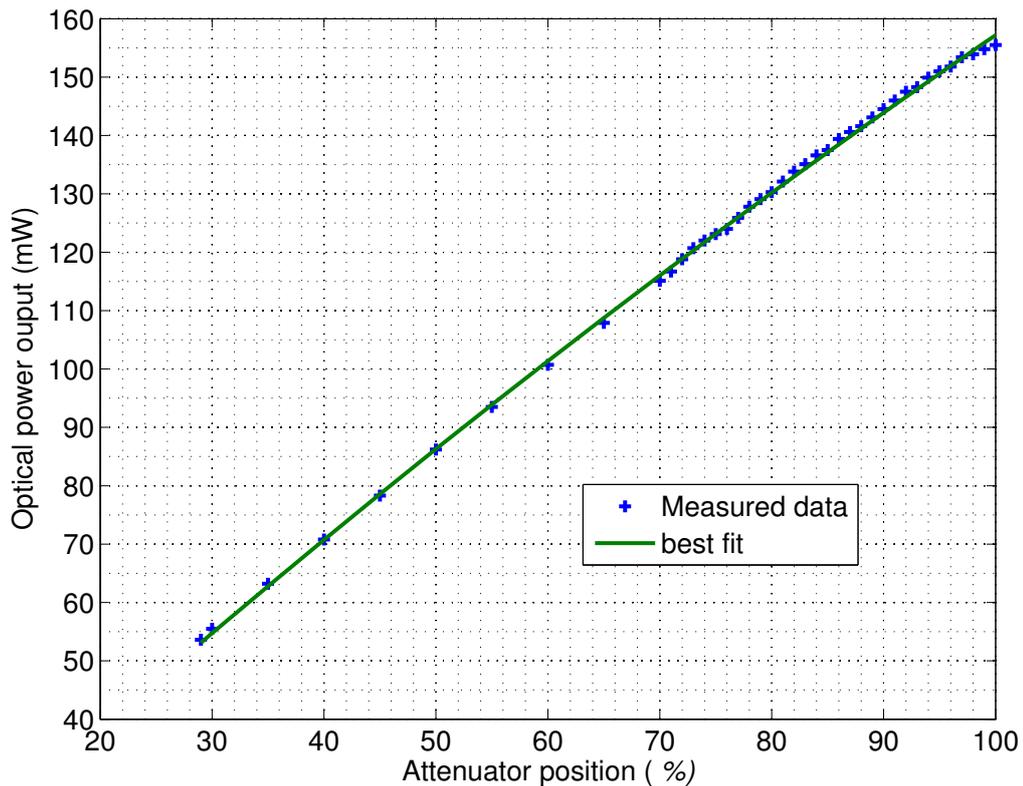


Figure 4.3: Measured *Low-Rep 1026 nm* power at the objective as function of relative polariser position.

4.1.3 High repetition rate system: 800 nm

This Ti:Sapphire chirped pulse oscillator system produces pulses with a spectral bandwidth of 30 nm centred at 794 nm at the rate of 11 MHz . The duration of the pulse is $50 \pm 17\text{ fs}$ (FWHM). Each pulse is produced with a 110 nJ of energy and deposits 70 nJ of energy on the target. The energy of the pulses was controlled using a pair of polarisers, as described above. A microscope objective with a NA of 0.8 was used for inscription. Detailed description of the *High-Rep 800 nm* system can be found elsewhere [140, 141].

4.1.4 Sample preparation and alignment

This section describes the procedure adopted to align the sample surface orthogonal to the laser beam axis on the *Low-Rep 800 nm fs* laser system. A slightly modified procedure was used on *Low-Rep 1026 nm fs* systems. Cleaning is essential to remove any impurities on the surface of the sample that could have the potential to diffract the light. Prior to mounting the samples to the stage, they were cleaned using acetone/methane with care.

There are two methods of fabricating the desired patterns. Either by scanning the beam whilst the sample is fixed or translating the sample relative to the beam. In all of our systems, the samples were mounted on a multi dimensional translational and rotational stage. Figure 4.4 shows an illustration of a sample mounted on the stage (*Low-Rep 800 nm* system). Focusing objective is fixed and the structures were inscribed by translating the sample along the $-y$ direction sample was translated in the negative axis direction. Position of the structure, below the surface was controlled by moving along $\pm x$ direction.

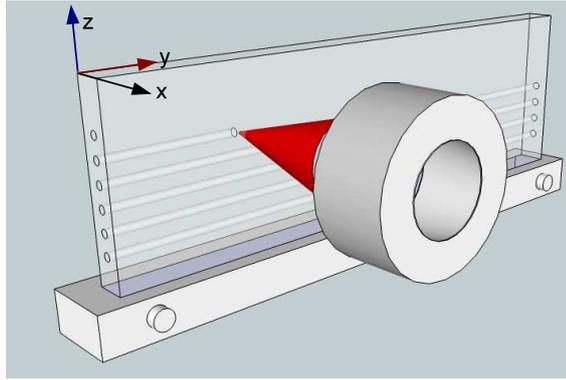


Figure 4.4: Example of sample alignment in Low-Rep 800 nm fs system.

Aligning the sample with respect to the laser focus is one of the critical factors in the fs inscription technique. The technique adopted here yields nearly $\pm 1 \mu\text{m}$ accuracy in the alignment. A planar sample can be aligned accurately perpendicular to the laser beam by measuring the position of the sample surface at (minimum) three locations.

The energy of the laser pulse was reduced substantially to obtain the diffraction at the surface of the sample without any material damage. However, the intensity is high enough to produce an observable diffraction pattern, when the focus hits the air-glass interface. The front surface of the sample was placed close to the objective and moved backwards slowly until the diffraction pattern from the front surface-air interface is observed in real-time using a CCD camera. The sample was then moved back and forth along the x-axis (beam axis) until the observed pattern size is minimised. The diffraction of the laser at the surface of the sample was used to determine the position of the surface at same three locations with respect to the laser beam. The relative positions of these locations were used to determine and correct the tilt of the surface.

Alignment is assessed by scanning the beam along the edges, with a low laser energy. Improper alignment will result in an observable change in the diffraction pattern as the sample is scanned along the vertical and horizontal directions. If the sample is perfectly aligned, one would expect to barely see any changes in the pattern. If noticeable changes are observed in the diffraction pattern, then the

alignment procedure was repeated, until the changes become unobservable.

The alignment procedure adapted for *Low-Rep 1026 nm* is slightly different, since this system uses an off-axis camera for observation. Laser pulses with substantially reduced power were fired at the sample to identify the surface where it produces plasma observable by naked eye. Following this, laser pulses with energy slightly above the ablation threshold were fired to produce surface ablation. These defects were examined using the off-axis camera. This procedure allows to identify the exact position of the surface, and to estimate the sample tilt. The goniometer was finely adjusted to minimise the tilt.

4.1.5 Writing direction

All of the test patterns presented in this thesis were designed and coded so that during the inscription process the structures at the lowest level (i.e. furthest from the focusing objective) were fabricated first in order to avoid scattering from the modified regions whilst inscribing subsequent layers. Figure 4.5 illustrates an example of this 'bottom-to-top' inscription process.

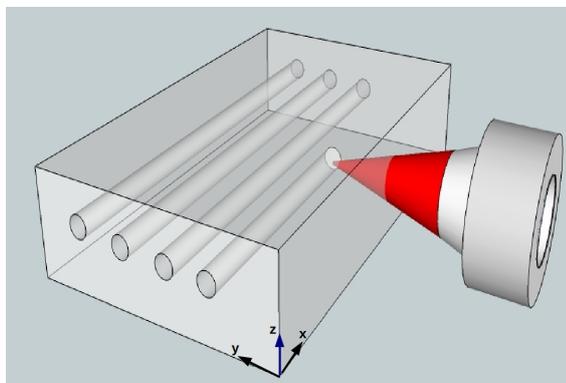


Figure 4.5: Example of writing direction (bottom-up) used for inscription of the OCT-phantoms.

4.1.6 Stage motion

Spacing between the modifications caused by individual pulses (i.e. points) along the inscription direction can be varied by varying the speed of the translation

stage. Each laser pulse produces material modification.

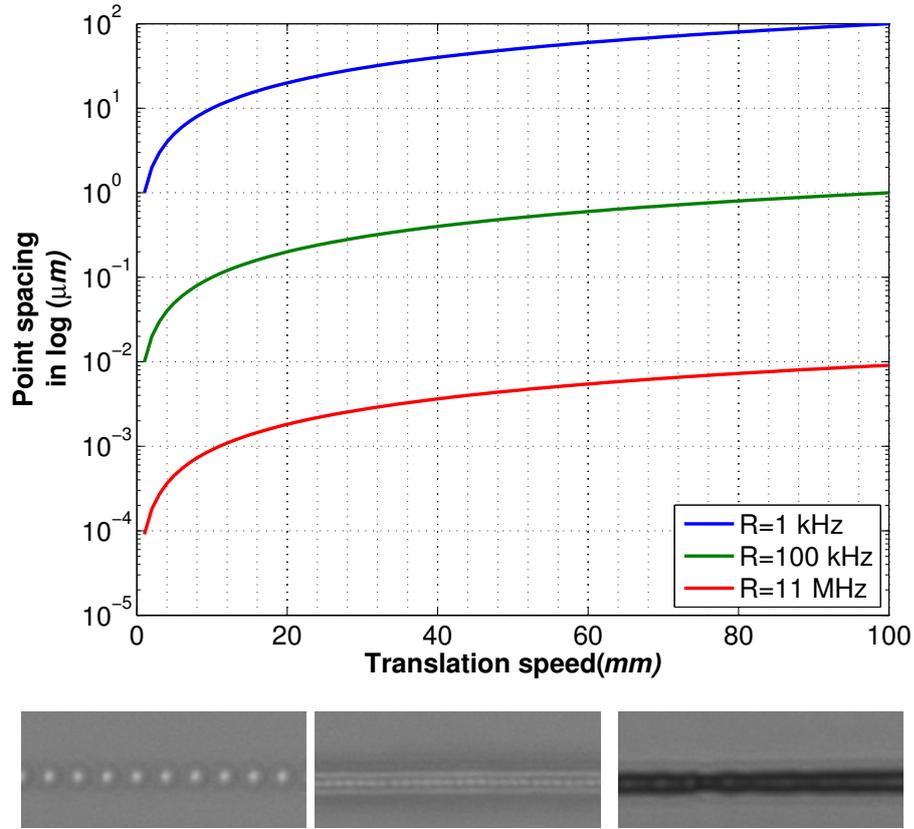


Figure 4.6: Point spacing as function of translation speed for three repetition rate (top). Microscope images of example structures inscribed by Low-Rep 800 nm (bottom). Microscope images shows structures inscribed a translation speed of 1 mm/s (bottom left), 0.6 mm/s (bottom middle) and 0.1 mm/s (bottom right). Structures were inscribed at a depth of 100 μm with 35.6 nJ energy.

Figure 4.6 show point spacing as function of translation speed (top) and microscope images of example structures produced by the *Low-Rep 800 nm* systems (bottom). 1 kHz, 100 kHz and 11 MHz at various translation speed. Microscope images show structures inscribed at a translation speed of 1 mm/s (bottom left), 0.6 mm/s (bottom middle) and 0.1 mm/s (bottom right). Structures were inscribed at a depth of 100 μm with 35.6 nJ energy. Individual points can be created with *Low-Rep 800 nm* system, if the sample is translated faster than 1 mm/s and assuming a spot size of ≈ 1 mm. The number of pulses deposited per unit area (flux) is also a key parameter that determines the quality of the produced structures.

4.2 Optical Coherence Tomography

A total of seven different OCT systems (Michelson Diagnostic OCT, Santec-OCT, two Thorlabs OCT systems, RT-OCT and two systems at Vienna) were used over the course of this study. The quantitative characterisation methods described in this thesis were applied only to Michelson Diagnostic OCT system, since this was the only system that was accessible over the entire period. Here onward, Michelson Diagnostic OCT system will be referred to as *MD-OCT system*. The system details and specifications of the *MD-OCT system* and Santec OCT systems are given below and the details of a home built OCT (RT-OCT) system are given in section 10.1. Table 4.2 summarises the properties of MD-OCT and Santec-OCT systems.

Table 4.2: Properties of the OCT systems

	MD-OCT	Santec-OCT	RT-OCT
Central wavelength λ (nm)	1305	1305	1325
Spectral width (FWHM) $\Delta\lambda$ (nm)	75	170	100
Axial PSF (FWHM) δ_z (μm)	11.4	8	7.7
Lateral PSF (FWHM) δ_{xy} (μm)	8.4 ± 2	23.1	not measured ⁵
imaging depth (mm)	1	3	not measured

4.2.1 Michelson diagnostic OCT (MD-OCT)

A SS-OCT system by Michelson Diagnostic, UK (EX1301)⁶ was used as the main OCT system for characterising the phantoms and developing the quantitative characterisation and calibration techniques presented in this thesis. In addition, accelerated GPU based data processing was also developed based on the data from MD-OCT.

This MD-OCT system is an SS-OCT system that uses a swept-source laser (HSL-2000, Santec, Japan) operating at a central wavelength of $\lambda = 1305 \text{ nm}$ with spectral width of $\delta\lambda = 75 \text{ nm}$ (FWHM) and an acquisition bandwidth of 150 nm .

⁶This OCT is in National Physical Laboratory (NPL), and have been used by P.H.Tomlins and P.D.Woolliams for their previous studies [1, 31, 120].

CHAPTER 4. EXPERIMENTAL SET-UP

The effective axial (FWHM) PSF of the system, estimated using the coherence length in equation (2.1.1) is $\delta_z = 11.4 \mu m$. The system uses an objective with a numerical aperture of $NA = 0.059 \pm 1 \times 10^{-3}$, that gives an expected lateral PSF of $\delta_{xy} = 8.4 \pm 2 \mu m$ estimated using the equation (2.1.1). Previous experimental measurements shows this particular OCT system has an axial and lateral resolutions of $10.9 \mu m$ and $8.2 \mu m$ respectively at the focal region and resolution degradation was observed away from the focal region [31].

The OCT system has an in-air axial and lateral (i.e. spacing between A-scans within a B-scan) pixel spacing of $4.16 \mu m$ and $5.61 \mu m$ respectively. This instrument comprises four interferometric channels focused at different depths ($250, 500, 750$ and $1000 \mu m$) to obtain an extended depth of field. The focal depth range of each channel is about $250 \mu m$ giving a total imaging depth of approximately $1 mm$. The results from a single channel was used throughout this chapter, unless explicitly stated otherwise. The corresponding maximum absolute sensitivity of the respective channels 1-4 are $\{93.7, 92.6, 90.1, 86.6\} dB$.

The scanning system comprises of a single galvanometer and a motorised linear translation stage (Z625B, Thorlabs Ltd, UK). The B-scans are produced by scanning the beam using the galvanometer at a software limited A-scan rate of $10 kHz$. To form the volumetric data (C-scan), multiple B-scans are obtained whilst translating the sample orthogonal to the galvanometer scan axis. The spacing between the B-scans can be arbitrary.

Figure 4.7 illustrates the galvanometer scan (red line) used to obtain B-scans and a linear stage that is used to obtain volumetric data. B-scans were obtained (approximately) orthogonal to the structures, which are indicated by the parallel lines on the the sample. Linear stage moves along the (positive) y-axis.

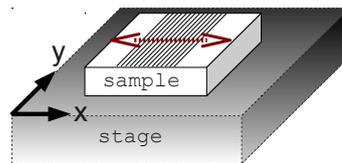


Figure 4.7: The galvanometer scan (red line) used to obtain B-scans and linear stage used to obtain volumetric data.

4.2.2 Santec OCT

The Santec OCT system (Santec, Japan) consists of a swept source spectral domain system that employs a light source, interferometer, portable scanning probe and computer acquisition hardware. This system has an axial and lateral resolution of $8 \mu m$ and $23.1 \mu m$ and has $3 mm$ imaging depth (i.e. $\pm 1.5 mm$ from DOF). The acquisition and processing of the data were customised for data throughput over the main board bus. A waveform digitiser (AlazarTech PCIe ATS-VI AT9462) with FPGA (Alazar Technologies Inc, Pointe-Claire, OC, Canada) were used for processing data, which reduces the processor overhead and stabilises the system performance.

The Mach-Zehnder interferometer (IVS-2000) was customised by adding an extra attenuation unit in the reference arm. A pair of galvo-mirrors were used for acquiring the volumetric data, with a maximum scan area of $15 mm \times 15 mm$, that takes around $0.5 s$. The system can produce approximately 30 B-scans (frames) per second, without any averaging or filtering. It is worth noting that the Santec-OCT data is available only in log-scale which was converted to linear scale using the equation below:

$$D_{lin} = 10^{0.1D_{db}}, \quad (4.2.1)$$

where D_{lin} and D_{db} refer to the data in linear and log scales.

4.2.3 Obtaining Data

Volumetric OCT data were obtained from both OCT systems. Raw intensity data were saved as 16-bit bigendian⁷ binary format by the MD-OCT. The volumetric data were processed using the CPU-GPU hybrid model described in section 9. This processing follows the general OCT data processing procedures: resampling from linear in wavelength to linear in wavenumber using second order Lagrangian

⁷Refers to how bytes are ordered within computer memory

interpolation, multiplication by Hann windowing function, and fast Fourier transform into spatial domain. Finally the modulus was computed. With the Santec OCT system, FPGAs are used for processing hence the processed data/images were saved as 32-bit floating point binary format.

4.2.4 Slicing angle

During data acquisition, samples were slightly tilted to avoid strong specular back reflection from the substrate surface. However this can distort the image. For example, a cross-section of a cylinder looks asymmetric and distorted in the image, when the scan direction is not orthogonal on its basis. Figure 4.8 illustrates a cylinder sliced at an angle θ (i.e. angle between structure and the beam). Line vector and unit normal vector are denoted by \vec{L} and \hat{n} (i.e. normal to B-scan).

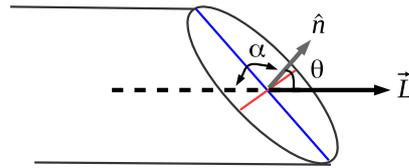


Figure 4.8: Example cross section when structures are scanned at an angle other than normal (i.e. wedge). The vectors \vec{L} , \hat{n} and θ denotes the line vector, unit normal vector and scan angle (i.e. angle between structure and beam) respectively.

The size of the observed ellipse is related to the angle of intersection θ between the scan-plane and the structure. Assuming the structures are cylindrically symmetric, the cross section of the structures looks like an ellipse when sectioned along an orientation other than normal to the lines. The semi-minor axis r_1 will remain the same as the structure radius r_r and semi-major axis r_2 is determined by the cutting angle α . Neglecting the effects of the optical aberration, cutting angle can be approximated by taking the position of the point on the first and the last frame.

CHAPTER 4. EXPERIMENTAL SET-UP

The area of the ellipse face is given by:

$$A = \pi r_r r_2 \cos(\alpha). \quad (4.2.2)$$

The line vector \vec{L} can be determined from the coordinate of the line at the first and last frame. A line passing through two points in three dimension (x_1, y_1, z_2) and (x_2, y_2, z_2) can be characterised by the parametric vector equation $x = \vec{x}_1 + (\vec{x}_2 - \vec{x}_1)t$ (or $x = x_2 - x_1$). The angle of intersection of this line with the plane (B-scan) can be calculated from the dot product

$$\hat{n} \cdot \vec{L} = |\hat{n}| |\vec{L}| \sin(\alpha). \quad (4.2.3)$$

The angle between the unit normal vector \hat{n} and the line vector \vec{L} can be calculated using the dot product. The semi-major axis of the sliced cross section can be computed by

$$r_2 = \frac{r_1}{\sin(\alpha)} = \frac{r_1}{\cos(\pi - \alpha)} = r_1 \sec(\theta), \quad (4.2.4)$$

where $\theta = \pi - \alpha$.

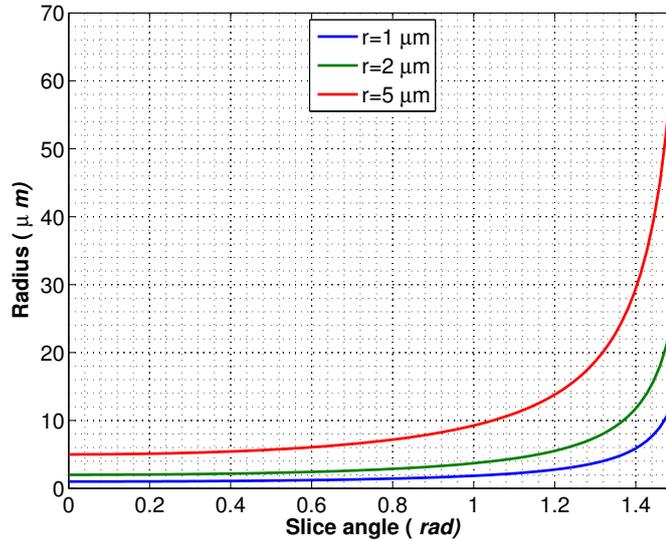


Figure 4.9: The radius of the cross section (semi-major axis) as a function of scan angle α .

Figure 4.9 shows the radius of the semi major axis versus the scan angle α , for a cylinder with semi minor axis of $5 \mu m$ and $2 \mu m$. The results shows that the effective elongation due to the tilt in the sample can be neglected, if the tilt angle is less than approximately $0.3(17.2)$ and $0.7(40.1)$ radians (degree) for the structures with width $5 \mu m$ and $2 \mu m$.

4.2.5 Point selection

A region of interest (ROI) was selected around the structures for further analysis such as estimating FWHM widths. Figure 4.10 shows the B-scan and the region of interest selected from the B-scan of the PSF pattern obtained by the MD-OCT system. A Matlab image processing toolbox was used to automatically identify the centroids and peak positions of the points in the ROI. Two copies of ROI data, both in linear scale, were used during the point identification process: the points were extracted from the “original” data that was unaltered, and the “duplicate” data was used for locating the coordinates (centroid and peak location) of the individual points.

The low frequency spatial information from the “*duplicate*” ROI was removed by applying an unsharp mask, that subtracts a strongly Gaussian blurred image from the original. The unsharp filter was created with a parameter of $\alpha = 0.2$ that controls the shape of the Laplacian, its values are given by the equation (4.2.5). An adaptive median filter of size 3×3 was also used to remove the background noise.

$$usf = \begin{bmatrix} -0.2308 & -0.5385 & -0.2308 \\ -0.5385 & 4.0769 & -0.5385 \\ -0.2308 & -0.5385 & -0.2308 \end{bmatrix} \quad (4.2.5)$$

CHAPTER 4. EXPERIMENTAL SET-UP

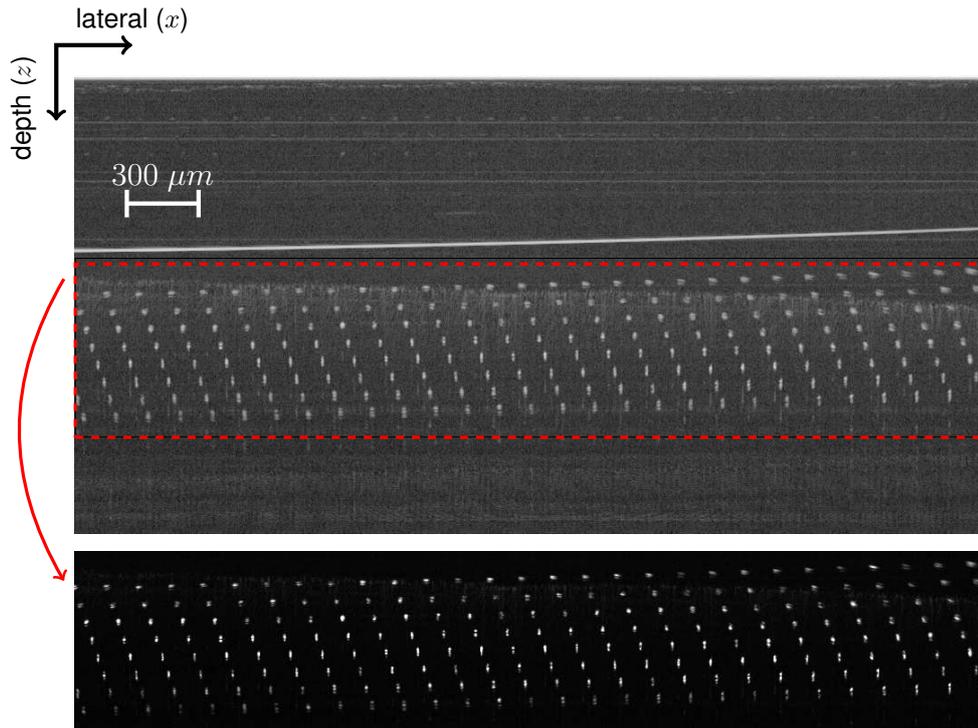


Figure 4.10: Step 1 of the point identification procedure: Extracting region of interest around the identified points. A B-scan (top) and cropped region of interest (bottom).

After applying a median filter, the resulting image was blurred using a two dimensional Gaussian blurring kernel. A Gaussian kernel of size 21×21 was created with a standard deviation $\sigma = 5$, as shown in Figure 4.11. Figure 4.12 shows an example of applying an unsharpened mask, median filter and Gaussian blurring to an ROI selected from the data shown in Figure 4.10 (i.e. B-scan from PSF pattern).

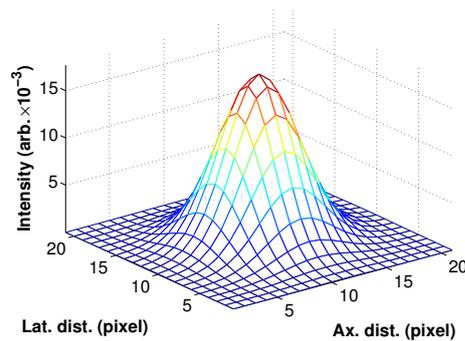


Figure 4.11: Example of a 2D Gaussian blurring kernel.

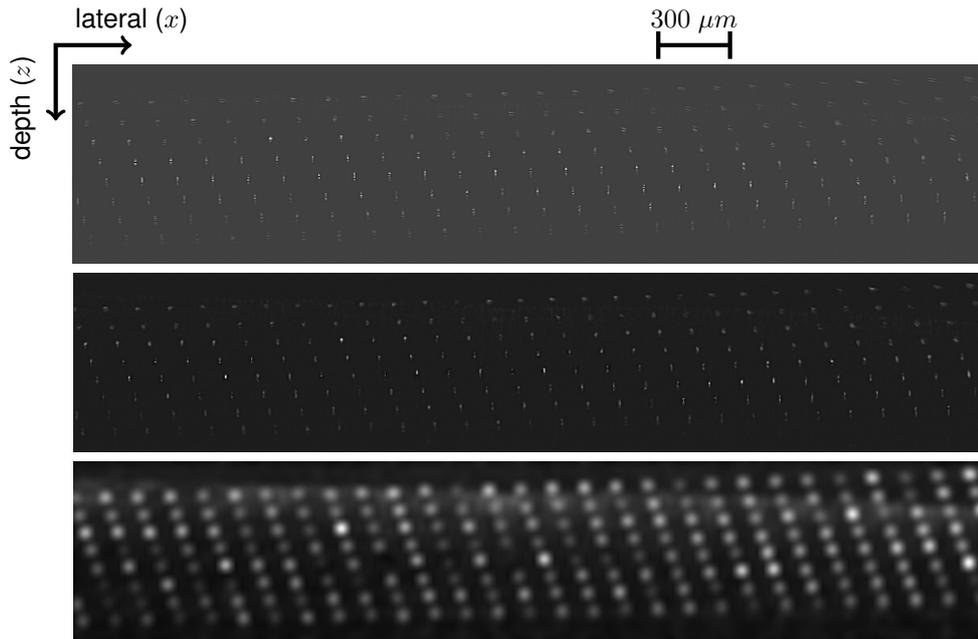


Figure 4.12: Step 2 of points identification procedure: enhancing image. B-scan after applying unsharpened mask (top), median filter (middle) and Gaussian blurring (bottom)

The resulting image was negated and the complements of the image was segmented using the watershed transform⁸. The centroid and centre position of the peak in each segment were computed. Figure 4.13 shows the selected points overlaid on top of the B-scan ROI.



Figure 4.13: Step 3 of points identification procedure: identifying the centroid of each points in the selected region. Red circles represent the centroid of each identified point.

Even after filtering the data, the persistence of some noise is unavoidable. This is clearly visible in Figure 4.13 where some red circles (e.g. centre top) represent the centroid data of noise data (i.e. there is no femtosecond inscribed structures in this regions). Therefore, after the points are identified, some point

⁸Matlab function

selection criteria were used to remove the noise points. Subsections with low signal to noise ratio were rejected by setting a threshold criteria. The peaks were rejected if there were less than six pixels below this threshold. Only those subsections satisfying these conditions are selected for fitting.

4.2.6 Selecting a sub-region

Because all the written structures in the sample are at discrete locations, this allows us to easily extract a subsection around each of the measurement point of interest. Subsections with low signal to noise ratio were rejected by setting a threshold criteria. The peaks with less than six pixels were rejected from the selection. Only those subsections satisfying these conditions are selected for fitting.

Using the coordinates estimated in the previous section a small area (dA) can be extracted around each point in the B-scan (ROI). Figure 4.14 shows an example of areas extracted from a B-scan ROI of the PSF pattern.

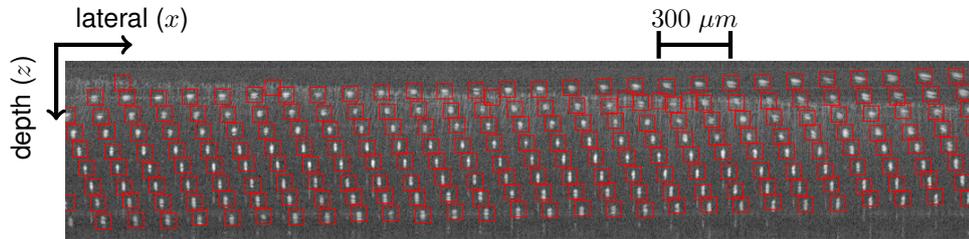


Figure 4.14: Regions selected for analysis. Red rectangle (21×21 pixels) represents the sub region to be selected around each point for quantitative analysis.

The coordinates of the points on the first frame of each sub-C-scan were computed. Using the centroid/peak position of each point, a small volume around $dV_{i,j,k}$ was extracted around each point.

A small volume dV of ($I \times J \times K$) pixels was extracted around each point, for calculating PSF at that point. The *en-face* and *side-view* are formed by summing the intensity values along the z and x directions respectively using the following

equations:

$$A_{ij} = \sum_k a_{ijk} \quad (4.2.6)$$

$$A_{jk} = \sum_i a_{ijk}$$

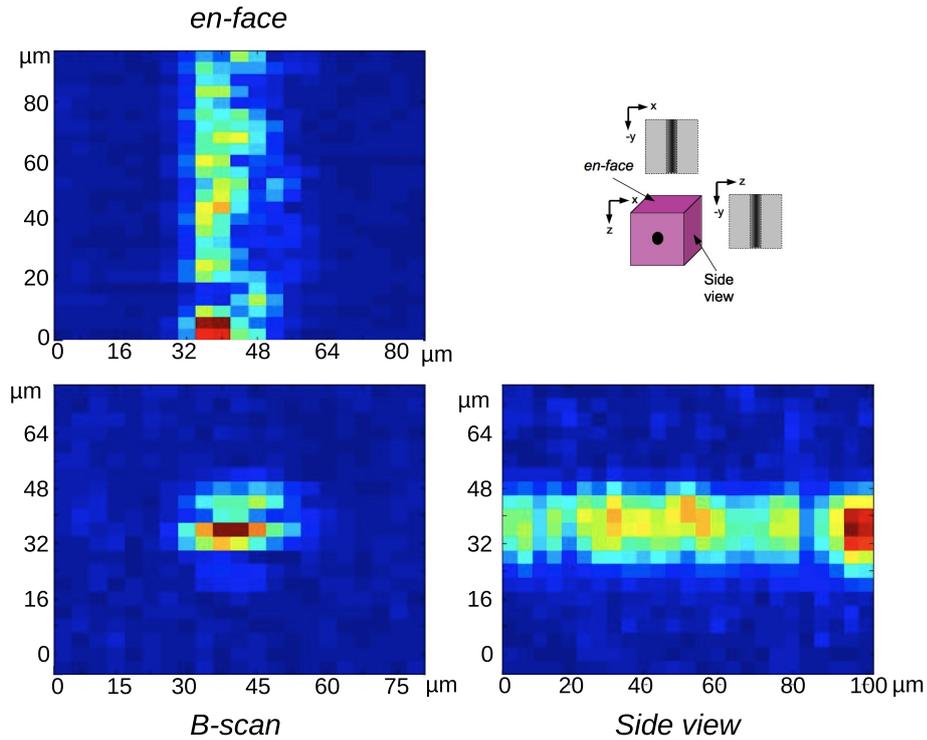


Figure 4.15: Example of a sub-section extracted around a structure. A schematic (top-right), cross-section (bottom left), *en-face* (top left) and *side-view* (bottom right) images of an arbitrary region. A $21 \times 21 \times 25$ pixels are extracted around a point, where each pixel is $4.1 \mu m$.

Figure 4.15 shows an example of a sub-section extracted around a structure. A schematic of the extracted sub section (top right), the cross-section (bottom left), *en-face* (top left) and *side-view* (bottom right) images of an arbitrary region. A $21 \times 21 \times 21$ pixels cube are extracted around a point, where each pixel is $4.1 \mu m$.

4.3 Conclusion

In conclusion, this chapter described the three fs laser systems and two of the three OCT systems used during this study. Description of the RT-OCT system is

CHAPTER 4. EXPERIMENTAL SET-UP

given in chapter 10. It has also described adjustments of the experimental set-up such as sample preparation and alignment for fs inscription, inscription direction. In addition to these, a procedure for identifying points (i.e. position and dimension of structures in B-Scan) was also given.

Chapter 5

Parameter Space Test

The quality of the inscribed micro-structures depends on a multitude of parameters. PST investigates effects on inscribed structures for different combination of inscription parameters. The key variables that influence properties of the inscribed structures are fluence, repetition rate, wavelength, speed and depth of inscription as well as the parameters of the objective used for focusing the beam inside the sample. Understanding the effect of all the variables is a time consuming task. Various teams around the world have been experimentally investigating these effects in relation to the induced material modifications as well as developing supporting theoretical models [45, 142–144]. This section describes the design of the PST phantoms in detail, including the selection of the optimal inscription parameters.

In this PST work, the main aim was to produce structures that are suitable for producing an OCT calibration and characterisation phantoms. The effect of the pulse energy and scan speed (i.e fluence) at various depths are investigated by inscribing a series of lines at different energy, scan speed and depth within the samples made of various materials. In addition, the suitability of different materials (fused silica, doped BK7, Borosilicate and PMMA) for making OCT phantoms were investigated.

The PST work was conducted on both low-repetition femtosecond laser sys-

tems. Table 5.1 shows the parameter ranges investigated during this study on both systems. An in-depth parameter space study on the *High-Rep 800 nm* was previously investigated by Dubov et al. with the aim of producing waveguides [140, 141].

The samples were then characterized using an OM equipped with QPM and an SS-OCT system. Some samples were characterised with a CM and a SEM.

Table 5.1: Summary of laser parameters used in the PST study in both low repetition femtosecond laser systems.

	Low-Rep 800	Low - Rep 1026
Energy (<i>mJ</i> per pulse)	16.89 - 80	12.53 - 155.5
Speed (<i>mm/s</i>)	0.1 - 2	0.1 - 20
Depth (μm)	0 - 775	0 - 625

5.1 *Low-Rep 1026 nm* system

5.1.1 Preliminary study

This section summarises the results reported by Tomlins et al. [1], which lays as the foundation for producing OCT characterisation and calibration phantoms using the femtosecond laser inscription technique. Here onward the fs inscribed phantom for calibration and characterisation of OCT system is referred to as OCT-phantom. Tomlins et al. published the initial work showing the potential of producing calibration phantoms for OCT systems using the *Low-Rep 1026 nm* system. A trial phantom was made by inscribing a three dimensional structure, that consists of series of lines, inside the bulk of a fused silica glass ($30\text{ mm} \times 25\text{ mm} \times 2\text{ mm}$, Ibsen Photonics, Denmark).

The PST phantom comprised a series of lines inscribed at different laser energy and depth, with *Low-Rep 1026 nm* system are shown in Figure 5.1¹. Lines

¹Samples were produced by Graham N. Smith using the *Low-Rep 1026 nm* system. Microscope characterisation of these samples were conducted by Janarthanan Rasakanthan. SEM, CM and phantom data were obtained by P.D. Woolliams and P.H. Tomlins at the NPL. Data analysis were done by Tomlins and Woolliams with a small contribution from Rasakanthan.

CHAPTER 5. PARAMETER SPACE TEST

were written in pairs at each energy level, with 100% lines written to facilitate location in microscopy and OCT images. This structure consists of 11 line pairs inscribed in 10 different layers, each at different depths, between 75 and 750 μm below surface, where layers are separated by 75 μm . The line pairs were inscribed with different energy values by focusing the laser beam using a $100\times/0.55\text{ NA}$ objective whilst translating the sample at constant speed of 1 mm/s . The line pairs were inscribed at 100, 80, 60, 40, 20, 10, 5, 2.5, 1, 0.5 and 0.25% of the total energy, that corresponds to 41.0 to 0.0125 μJ . The maximum laser power at the objective was 4.1 W (i.e. peak pulse power of 0.1 GW)².

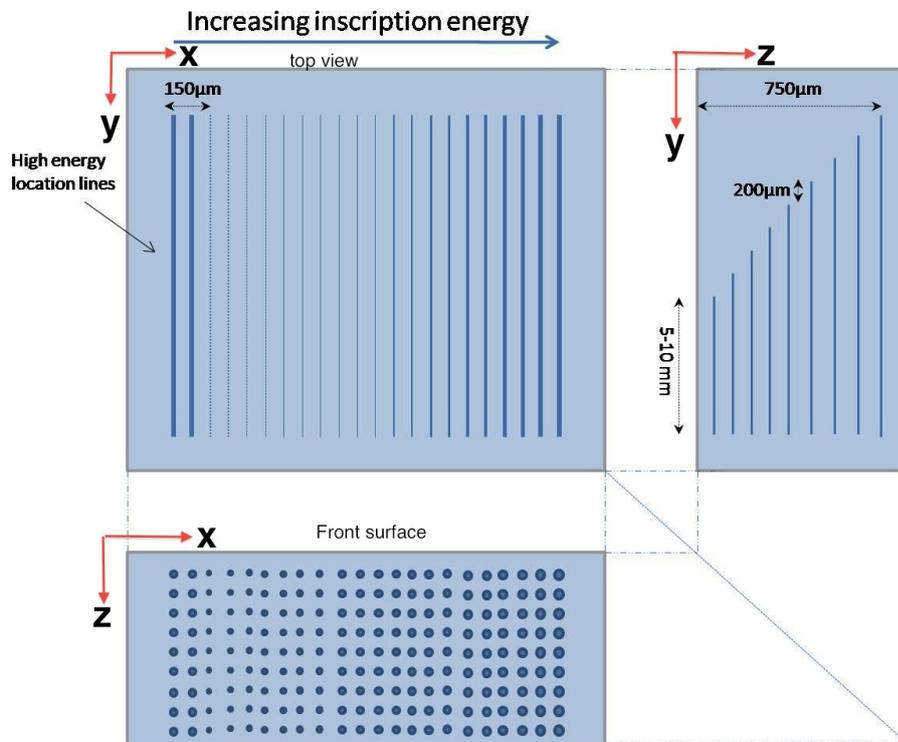


Figure 5.1: A schematic of the femtosecond inscribed OCT calibration phantom showing the paired lines of varying power increasing from left to right after initial location lines [1].

Firstly, the phantom was evaluated using an optical microscope (Axioskop 2, Carl Zeiss, Germany). A 20x magnification imaging lens (Plan Neofluar 20x/0.50) was used to acquire microscope images of the sample. The lateral width of the

²For the same $100\times/0.55$ objective used here, we noted that only about 5% of the power is transmitted through the objective and deposited onto the surface of the substrate.

size of the modified regions estimated for each of the inscription power and the different depths by measuring the inscribed line widths.

The variation of the defect size with the inscription energy is clearly visible in Figure 5.2a. Three line pairs are shown, where feature width increases with increasing inscription energy. The accumulative phase retardants of the inscribed structures were measured using QPM microscopy, from which the inscription was identified as a distinct phase modulation, as shown in Figure 5.2b. Results shows that the damaged regions have an induced phase difference, both positive and negative, compared to the undamaged regions.

Figure 5.3 shows the lateral size variation of the structure inscribed at $75 \mu\text{m}$ with varying energy, as measured using an optical microscope. Along any given line six measurements were taken and then averaged. Structures inscribed between 5 to 100% of the laser energy were visible under microscope, and had the lateral dimension between 0.5 and $7.5 \mu\text{m}$. Those structures inscribed with less than 5% energy were not easily visible (they could be distinguished but not reliably measured) by the optical microscope.

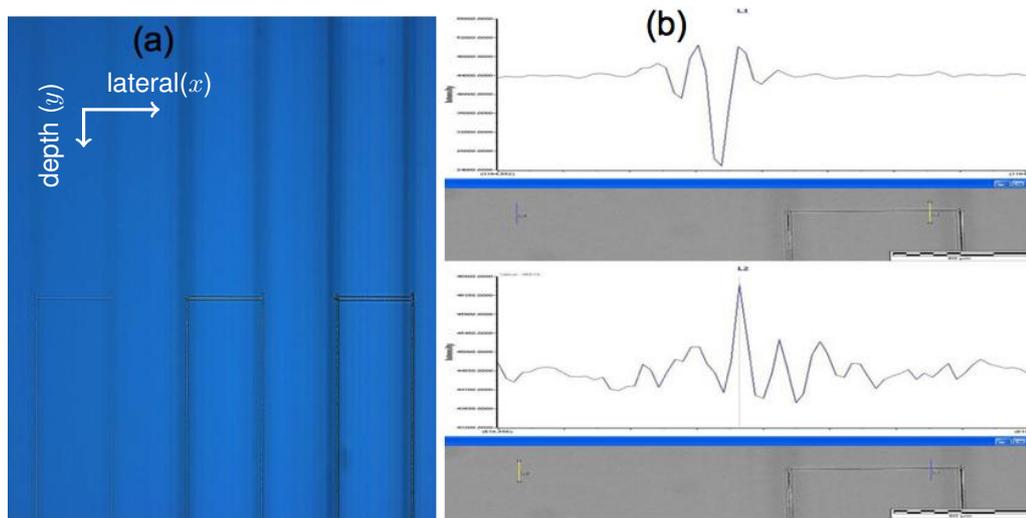


Figure 5.2: a) Microscope image of three pairs of lines showing a variation in the feature width with varying inscription power b) QPM measurements of two lines showing the contrast between above material threshold void creation (above) and below material threshold index change. The yellow line in the grey image marks the measurement area for relative phase plot [1].³

³Images were obtained by G.N. Smith.

CHAPTER 5. PARAMETER SPACE TEST

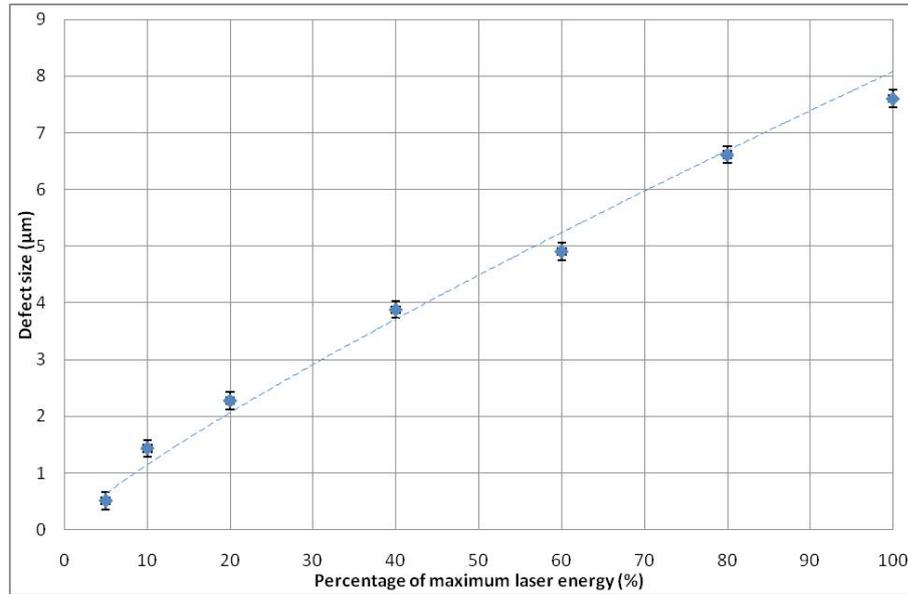


Figure 5.3: Variation of measured apparent defect size with laser power setting used taken at a depth of $75 \mu\text{m}$ from the surface, as measured using an optical microscope [1].⁴

The test phantom was also evaluated using an SS-OCT system (*MD-OCT system*). Figure 5.4 shows the OCT cross-section of the structure, obtained by focusing approximately $150 \mu\text{m}$ below the sample surface. Results were averaged over 10 B-scans and presented in logarithmic intensity scale without refractive index correction for fused silica substrate. Lateral and axial directions are denoted x and z respectively. Image was cropped around the damages produced by fs inscription, cross-sections of the modified regions are shown as bright points in the B-scans. The lines inscribed at 100% of the laser power are at the far edges with the inscription power decreasing from the left hand side (columns 1 – 20). Columns 21 and 22 (far right) represent lines written at 100% power for identification. Points visibility near the focal region extends to column 16, which corresponds to 2.5% of the inscription energy. This is the smallest inscribed structure that is visible under OCT.

⁴Data were obtained by myself.

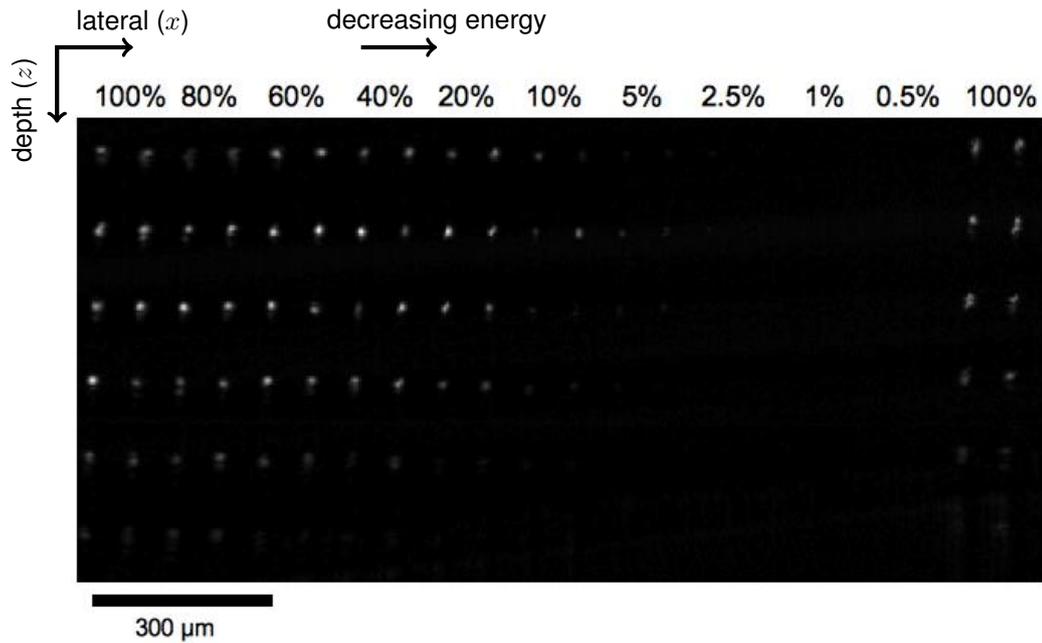


Figure 5.4: OCT cross-section of an engraved sample, scaled in logarithmic intensity. Lateral and axial dimensions are denoted as x and z respectively. Femtosecond inscription is denoted by the bright points. Rows and columns were written at $75 \mu\text{m}$ intervals. At each depth, pairs of lines were written at laser powers varying from 0.5 to 100%. The OCT image has not been corrected for the silica substrate refractive index [1].⁵

To determine the optimum etching time, another trial sample was fabricated, which contained lines inscribed at a constant energy and at a single speed, at a depth of $100 \mu\text{m}$ below surface. Sample was cleaved transverse to the inscribed lines, into small pieces and immersed in hydrofluoric (HF) acid (40% concentration acid was diluted 20 : 1 in distilled water) for varies duration. Etched sample were then cleaned and examined using a CM. Figure 5.5 shows images obtained with a LEXT CM system, after various etching time. The hole region in the sample are shown as noise in the data due to no detected signal being interpreted⁷.

⁵Image were obtained by P.H. Tomlins and P.D. Woolliams.

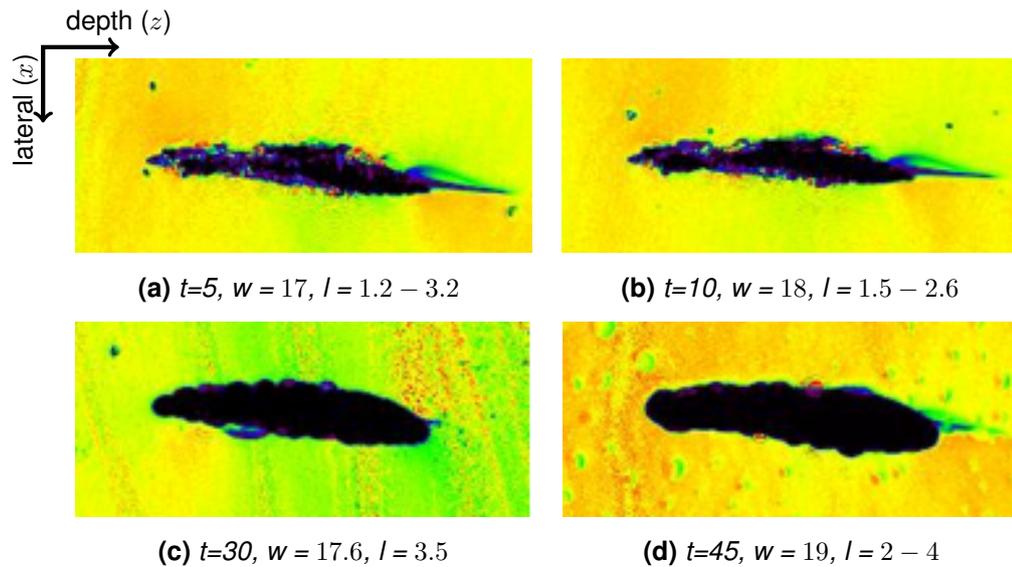


Figure 5.5: CM images of the cross section of modified regions after etching in HF acid for various times, where t , w and l are the etching time in minutes, width and length of the structures in μm respectively ⁶.

The sample was partially cut through the back side and then cleaved and etched to reveal the cross sectional profile of the damaged regions. Figure 5.6a and 5.6b show the CM image of a small regions, and SEM image of the cleaved face that reveals morphology of the cross section of the damaged regions⁷. CM image shows three line pairs inscribed with highest energy ($E = 100, 80$ and 60% of the maximum energy) at six different depths. SEM image presented here is rotated by 90° (i.e. inscription laser is not pointing from left to right). Axis of inscription and depth are denoted by x (i.e. along columns) and variation of energy is denoted by z (i.e. along rows). CM and SEM images reveals the asymmetry of the inscribed structures. This is an expected behaviour intrinsic to the lower repetition rate of the laser system used in this study. Additional optical components such as a cylindrical lens or slits can be used to reduce this asymmetry [81].

⁶Images were obtained by P.D. Woolliams.

⁷Data were obtained by Peter D Woolliams at NPL from the trial samples produced by Graham N Smith using *Low-Rep 1026 nm* laser system.

CHAPTER 5. PARAMETER SPACE TEST

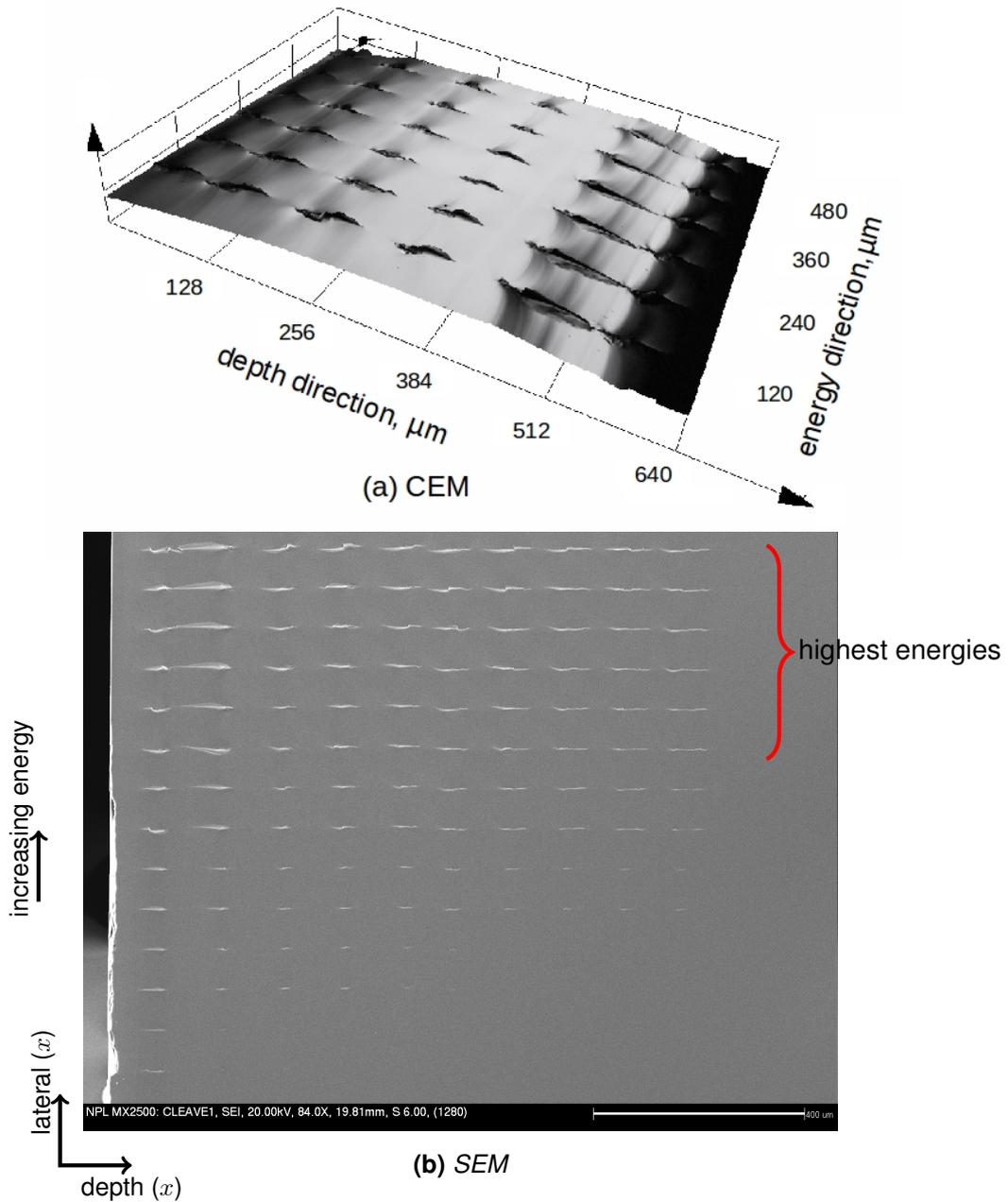


Figure 5.6: *CM (top) and SEM (bottom) images of the cross section of the cleaved and chemically etched structures inscribed in fused silica sample. CM image shows the cross-sections of three pairs of structures inscribed at highest energies (i.e. top 6 rows in SEM image) and at 6 different depths. The SEM image shows the cross-section of all damages and it is rotated by 90°. Axis of inscription and depth are denoted by x (i.e. along columns) and variation of energy is denoted by z (i.e. along rows).*

Figure 5.7 shows the axial and the lateral widths measurements of the lines

CHAPTER 5. PARAMETER SPACE TEST

from phantom, obtained with a CM, where the structures were inscribed at $75 \mu\text{m}$ for different energy values (of the lines from the end face (etched surface)). The lateral diameters measured using CM are smaller shown in Figure 5.3, this could probably be due to the high contrast ratio of the etched sample making it easier to focus on the defect.

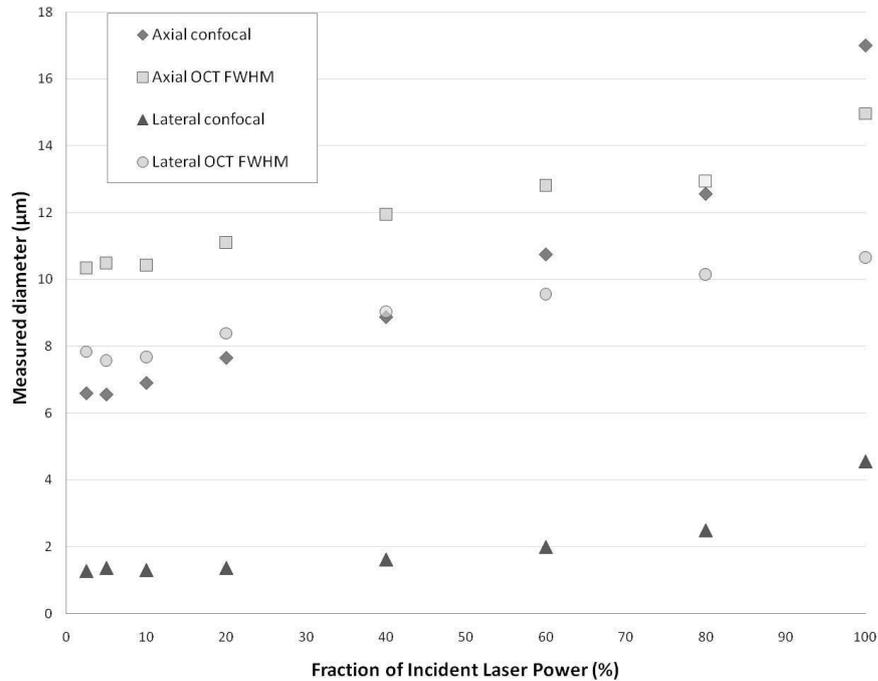


Figure 5.7: Axial and lateral width of the structures measured using an OCT and a CM [1].

Figure 5.8 shows CM images (grey) of the cleaved, etched phantom with corresponding OCT cross sections (colour) shown on an intensity linear scale. The OCT were obtained with MD-OCT system. images CM images were obtained after cleaving and chemically etching the sample. Images a-e correspond to 100% power, f-j to 20% power and k-o 10% power. Rows 1 – 4 correspond to depths of $75, 150, 225$ and $300 \mu\text{m}$ from the phantom surface. It should be noted that the full scale (lateral and axial) for the CM images is $75 \mu\text{m} \times 75 \mu\text{m}$ but only $40 \mu\text{m} \times \text{mm}$ for the phantom measurements. A detailed explanation of the oversampling procedure used here is given in section §3.7. The results show the clear trend in the size of the inscribed structure with pulse energy. There is also a notable change

CHAPTER 5. PARAMETER SPACE TEST

in the OCT measurement for points at different depths. The 10% laser power gives noticeable smaller features than the 100% laser power.

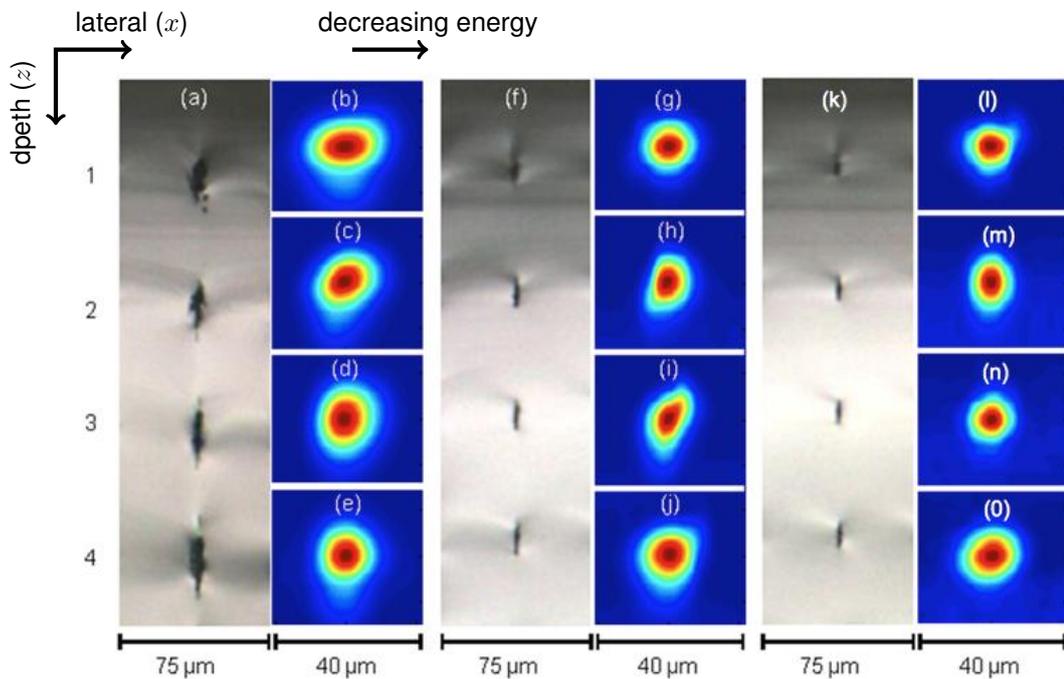


Figure 5.8: CM images (grey) of the cleaved, etched phantom with corresponding OCT cross sections (colour) shown on a intensity linear scale. Images a-e correspond to 100% power, f-j to 20% power and k-o 10% power. Rows 1 – 4 correspond to depths of 75, 150, 225 and 300 μm from the phantom surface. It should be noted that the full lateral scale for the CM images is 75 $\mu\text{m} \times 75 \mu\text{m}$ but only 40 $\mu\text{m} \times 40 \mu\text{m}$ for the phantom measurements [1].

The Michelson Diagnostic OCT was unable to resolve the structures inscribed at or below 2.5% the total power. Figure 5.9 shows the CM images of the cleaved and etched cross section of the structures inscribed using 2.5% of the total power at a depth of 75 to 300 μm below the surface. Imaged a-d corresponds with depths from the phantom surface of 75, 150, 225 and 300 μm below surface respectively.

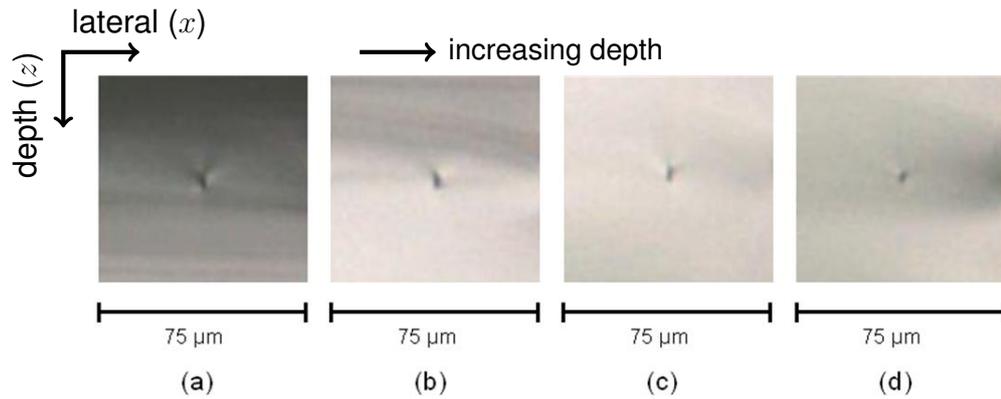


Figure 5.9: CM images of the cleaved and etched cross-section of the inscribed structures using 2.5% of the total energy. Imaged a-d corresponds with depths from the phantom surface of 75, 150, 225 and 300 μm below surface respectively [1].

The CM images, after chemical etching, show presence of elongation in structures inscribed at 10% power as well (Figure 5.8k). However, the effect is negligible in the OCT B-scan (Figure 5.8l-o). The structures inscribed at 2.5% power showed relatively small elongation (Figure 5.9). However, these structures are not reliably visible in OCT B-scans. These results imply that more work should be done between 2.5% and 10% laser power to optimize the line shape, size and reflectivity.

Discussion

All inscribed structures were visible under the optical microscopes. However, none of the structures inscribed for less than 5% of the laser power were reliably visible under OCT system. This suggests, that either the induced refractive index is smaller than the sensitivity threshold of the OCT system and / or inscription energy is much lower than the threshold energy required for the refractive index modification.

Microscope measurements show, that the lateral size of the structures (i.e. modified regions) are significantly smaller than the lateral resolution of the OCT system. Therefore, these structures can be approximated as delta functions, when used in the convolution of the object and instrument PSF to form the fi-

nal image. The structures inscribed at lower energy are nearly symmetrical. This allows simultaneous measurement of the axial and lateral PSF of an OCT system.

Axial bifurcation is observed in individual OCT B-scan images, for structures inscribed at energy levels 20% and above the threshold power. Bifurcation disappears in averaged image, which could indicate this phenomenon arises due to the interference between light reflected from the top and bottom of the structure that exhibit a noticeable depth elongation (Figure 5.8 a, f and k). Chemical etching results show that structures inscribed at higher energies have prolonged elongation along axial dimension (Figure 5.6).

5.1.2 Extended parameter space study

In order to optimise laser inscription, parameters suitable for fabrication of OCT phantom, wider range of speed, energy and depths were investigated. Lines were inscribed in a $30 \times 25 \times 2 \text{ mm}$ pure silica substrate (Ibsen Photonics, Denmark) by focusing the laser using a Mitutoyo M Plan Apo NIR $\times 100$ lens having a numerical aperture of 0.50 and a focal length $f = 2 \text{ mm}$ that yields a Rayleigh criterion focused spot size of $\approx 2.5 \mu\text{m}$.

The phantom used for extended PST comprised of series of lines inscribed at different depths, with different laser energy and translation speed as shown in Figure 5.10. Structures were inscribed at 10 different depths using five different speeds denoted respectively in columns $D_1 - D_{10}$ and rows $S_1 - S_5$. Each set (i.e. (D_i, S_j) , where $i = 1 - 10$ and $j = 1 - 5$) contains 20 lines inscribed at various energies, shown in Table 5.2 and the line patch within the set was set to $150 \mu\text{m}$. The columns were inscribed between $0.1 - 1.0 \text{ mm}$ below the surface, starting at $100 \mu\text{m}$ (left most) and each subsequent column was written $100 \mu\text{m}$ below the previous one. Rows S_1 to S_5 were inscribed by translating the sample at 1, 2, 5, 10 and 20 mm/s . First line within each set was inscribed at 152.42 mW laser power and used as markers to identify the lines inscribed with lower energies.

CHAPTER 5. PARAMETER SPACE TEST

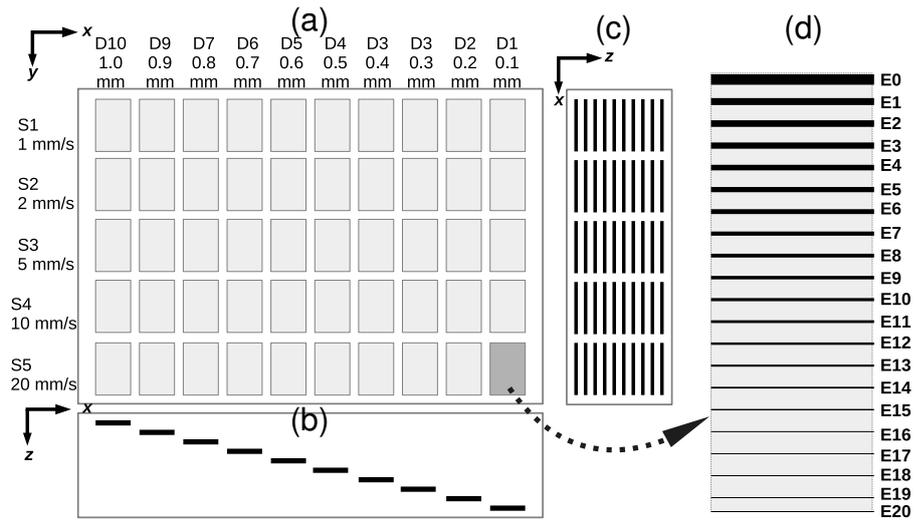


Figure 5.10: A schematics of the extended parameter space sample. Each of the gray region in (a) represents lines inscribed at 20 different energies as shown in (d).

Table 5.2: Line number in the phantom and their corresponding inscription energy.

Line number	Polariser rotation (%)	Actual energy (J)
1	100.0	152.4160
2	40.0	69.9760
3	35.0	62.2122
4	30.0	54.3110
5	25.0	46.2722
6	20.0	38.0960
7	15.0	29.7822
8	14.0	28.1030
9	13.0	26.4183
10	12.0	24.7280
11	11.0	23.0322
12	10.0	21.3310
13	9.0	19.6243
14	8.0	17.9120
15	7.0	16.1942
16	6.0	14.4710
17	5.0	12.7423
18	4.0	11.0080
19	3.0	9.2682
20	2.0	7.5230
21	1.0	5.7722
22	100.0	152.4160

CHAPTER 5. PARAMETER SPACE TEST

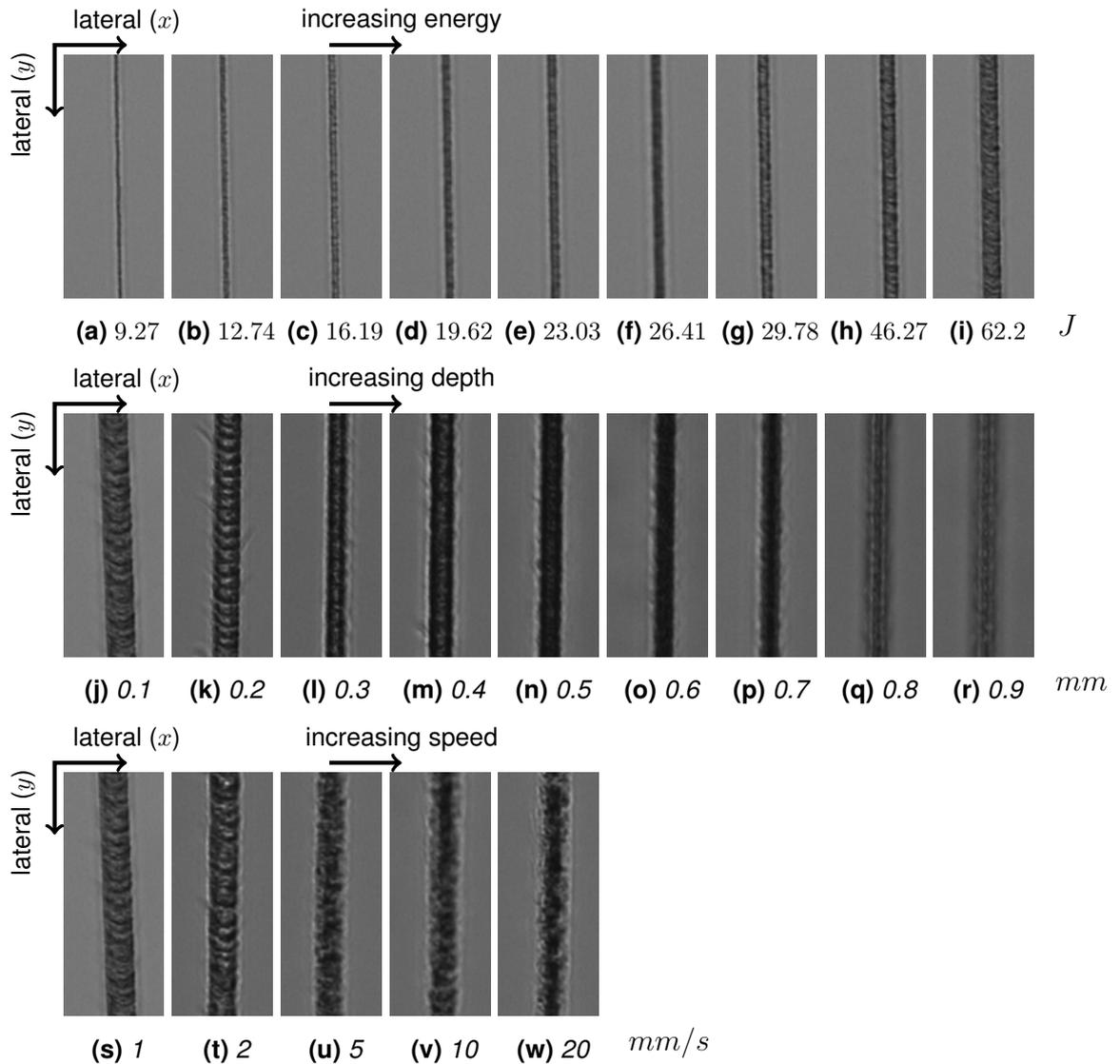


Figure 5.11: Microscope images of the structures showing variation in the feature width with varying energy, varying speed and varying depth. First row shows structures inscribed with 9.27 – 62.21 J laser energy at 1 mm/s speed at a depth of 100 μm below the surface. Middle row shows structures inscribed between 100 μm and 900 μm below the surface, with 40 J of laser energy and 1 mm/s speed. Bottom row shows structures inscribed with 69.98 J energy at 100 μm below surface speed between 1 and 20 mm/s. Y-axis represents the sample translation direction.

The phantom was first evaluated using an optical microscope (Axioscope 2, Carl Zeiss, Germany). A 20 \times magnification lens (Plan Neofluar 20 \times /NA0.5) was used to acquire images of the sample. The lateral width of the size of the modified regions were estimated from the microscope images for each of the inscription power, different depth and different speed by measuring the inscribed line widths.

CHAPTER 5. PARAMETER SPACE TEST

Along any given line, six measurements were taken and then averaged. The effect of lateral width variation of the inscribed structures with varying inscription power, varying speed and varying depth are clearly visible in Figures 5.11. First row shows structures inscribed with $9.27 - 62.21 J$ laser energy at $1 mm/s$ of speed at a depth of $100 \mu m$ below the surface. Middle row shows structures inscribed between $100 \mu m$ and $900 \mu m$ below the surface, with $40 J$ of laser energy and $1 mm/s$ speed. Bottom row shows structures inscribed with $69.98 J$ energy at $100 \mu m$ below surface speed between 1 and $20 mm/s$. Figure 5.12 shows microscope lateral widths of the structures obtained from the microscope images, for varying energy, varying speed and varying depth.

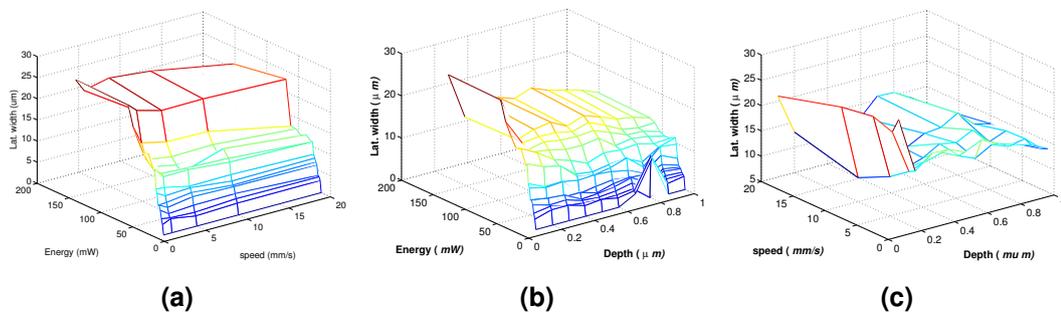


Figure 5.12: Variation of apparent lateral size for the structures as varying laser energy, speed and depth. Images show the average lateral width as function of a) speed and energy, obtained at $100 \mu m$ depth; b) energy and depth, inscribed with constant speed of $1 mm/s$; and c) speed and depth inscribed at constant energy of E .

Figure 5.13 shows the OCT cross-section of the sample in logarithmic intensity scale. The strongest lines, inscribed at 100% laser energy for identification, are at the far edges (i.e. column 1 and 22) with an inscription energy decreasing from left to right. Rows a-e and f-j were inscribed at $100 \mu m$ and $400 \mu m$ below the surface, whilst a-e (and f-j) were inscribed at a speed of $1, 2, 5, 10$ and $20 mm/s$.

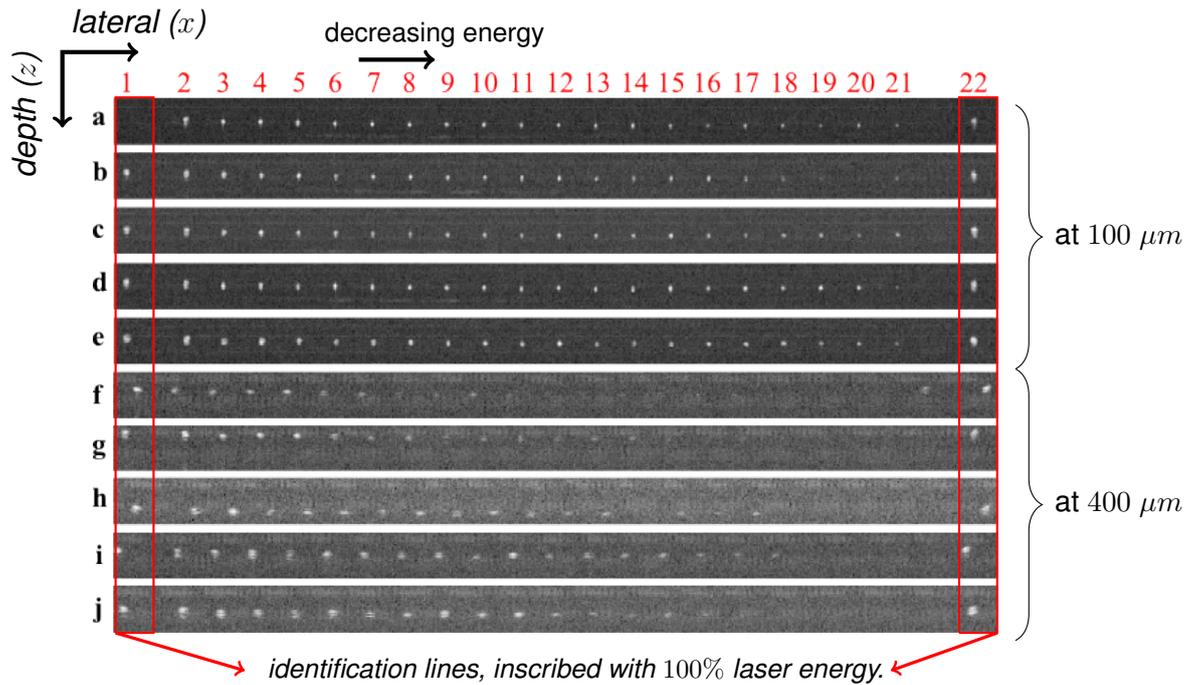


Figure 5.13: OCT cross-section of the sample in logarithmic logarithmic scale. Rows a-e and f-j were inscribed at $100\ \mu\text{m}$ and $400\ \mu\text{m}$ below the surface, whilst a-e (and f-j) were inscribed at a speed of 1, 2, 5, 10 and 20 mm/s.

Axial and lateral widths of the inscribed structures were estimated from the FWHM of a two-dimensional Gaussian function fitted to each detected point on OCT data. The results of these values averaged over 10 sequential B-scan slices are shown in Figure 5.14 at a depth of $100\ \mu\text{m}$. All structures were inscribed $100\ \mu\text{m}$ below the surface with a scan speed of 1mm/s and with a varying percentage of laser power of 100, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 3, 2, 1, where 100% corresponded to power $E_T = 157.17\text{mW}$. The measured axial and lateral widths were plotted as function of inscription energy and speed. A similar trend in variation of size is observed for different depths as well.

CHAPTER 5. PARAMETER SPACE TEST

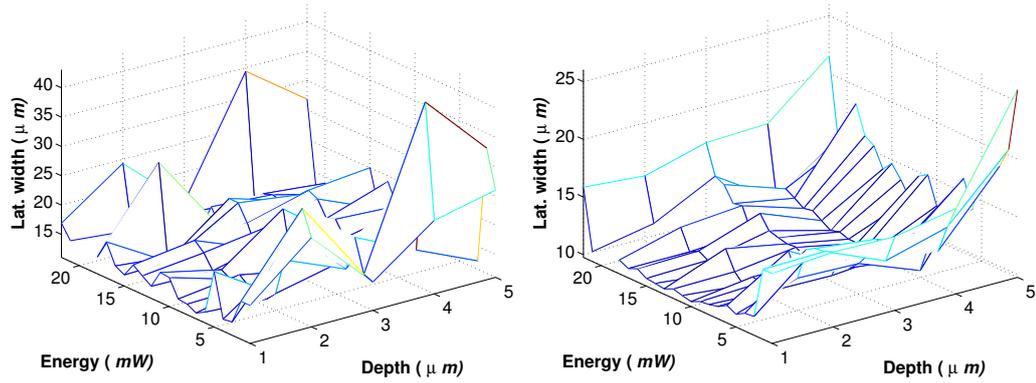


Figure 5.14: The FWHM axial (left) and lateral (right) width of the damaged cross-sections. Data were estimated by fitting a 2D Gaussian function to the OCT measurement. All structures were inscribed $100 \mu m$ below the surface with a scan speed of $1 mm/s$ and with a varying percentage of laser power of 100, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 3, 2, 1 percentage from a total power $E_T = 157.17 mW$.

Discussion

All inscribed structures, except the ones inscribed at $5.7 J$ (i.e. 22nd line), were visible in the microscope image. This suggests, that an inscription energy $> 5.7 J$ might be needed for any material modification in fused silica by *Low-Rep 1026 nm* system. Visibility of the structures decreases with depth in the OCT image and structures inscribed at lower energies are not visible at all. The size of the feature decrease with decreasing energy and stays the same with speed. This could suggest that the induced refractive index change is much lower than the sensitivity of the OCT system. Microscope measurements shows that size of the structures inscribed with energy below $< 24.73 J$ are much smaller than the lateral resolution of the OCT system. Therefore, these can be used to measure the PSF of the OCT system. Structures inscribed with an energy $< 38 J$ show axial asymmetry in the B-scan. Axial bifurcation is observable for structures inscribed at deeper depths. Microscope images show that the size and quality decreases slightly with depth.

The microscope and OCT results show that the lateral size increases with increasing energy at constant speed for all depths. The results indicate that the lateral size of the lines is much larger at lower speed ($1 mm/s$) than the struc-

tures inscribed at higher speeds, which have similar size for constant energy and depths. However, the quality of the structures inscribed at $1 - 2 \text{ mm/s}$ is much better than those inscribed at higher speed.

The lateral size of the inscribed structures decreases with increasing speed at higher energies. The structures inscribed at lower speed, at constant energy and constant depth are very similar. The lateral and axial size of the structures decrease with increasing depths, at higher energies probably due to absorption. Structures inscribed at different depths for constant energy and constant speeds are very similar.

The structures inscribed with 1% laser energy were not visible either using a OM or an OCT.

5.1.3 Material differences

The PST was conducted on three different materials, (anti reflection (AR) coated) fused silica, borosilicate and FeO_2 nano-particle embedded polyurethane phantoms⁸. This section describes the design of the PST samples in details. The produced structures were characterised using an OM and several OCT systems. This section presents the parameter study conducted on these materials.

In previous designs (Figure 5.10), a volumetric OCT data collection was required for each set given a total of 50 OCT volumes. The sample contains patterns inscribed at 10 depths, 5 speeds 20 different inscription energies. A B-scan (or volume) is required only take image of 1 depth, 1 speed and all energies. Therefore in order to evaluate all of the inscribed structures, we need 50 B-scans or volumes. This determines an unnecessary bottleneck for the OCT data acqui-

⁸This work was carried out in collaboration with M. Sandrian, P.H.Tomlins, P.D.Woolliams and G.C.B.Lee. PST patterns were designed and coded by myself with G.C.B.Lee contributing to debugging the G-code. The FeO_2 embedded polyurethane samples (unpolished surface) used in this study were fabricated by M.Sandrian, P.H.Tomlins and P.D.Woolliams. In addition to this, at a latter date, we have inscribed the PST pattern inside few more polyurethane samples with varying particle density. These samples were made by P.D.Woolliams and P.H.Tomlins as resolution phantom and the surface was polished to optical quality. The details of fabricating these phantom can be found elsewhere [16].

Table 5.3: Details of the parameter space phantom study using Low-Rep 1026 nm laser system.

	Materials		
	Fused silica (AR coated)	Borosilicate	Nano particle embedded resin
Below surface (μm)	100	100	100
Line spacing (μm)	150	150	150
Layer spacing (μm)	50	50	50
Layer offset (μm)	25	25	25
Line length (μm)	1000	1000	1000
Energy (nJ)	1 – 100%	5 – 100%	5 – 100%
Depth (μm)	0 – 1000	0 – 450	0 – 1000
Speed (mm/s)	0.1 – 20	1 – 20	1 – 20
Objective	100 \times /0.55 NA	100 \times /0.55 NA	100 \times /0.55 NA

sition, processing and analysis. Therefore, the design of the parameter space test was slightly altered to write a stack of lines at different depths as a preliminary test where each layer is off-set slightly. The same design as Sensitivity pattern presented in Section §6.2.3 was used for PST. This modified design significantly simplified the OCT characterisation process.

Each PST was inscribed with varying energy and constant columns were inscribed with a different speed. Each subsequent layer, in depth, was separated by $75 \mu m$ and offset by $25 \mu m$ in horizontal plan (i.e. $+x$ direction).

Figure 5.15 shows a photograph of a PST sample and a schematic representation of nano-particle embedded FeO_2 phantom⁹. In the sample shown here, each columns were inscribed (from left to right) with $\{15, 10, 5, 12, 8, 4, 3, 2, 0.1\}mm/s$ respectively. Each pattern consists of 25 pairs of line, inscribed between $42 - 830 nJ$ at 12 different depths between $100 \mu m$ and $1 mm$. Similar patterns were inscribed in other materials. Additionally, the energies and speeds were also varied to investigate the parameter space.

⁹These sample was made by M.Sandrian, P.H.Tomlins and P.D.Woolliams. The parameter space test pattern were inscribed by G.C.B.Lee and my self.

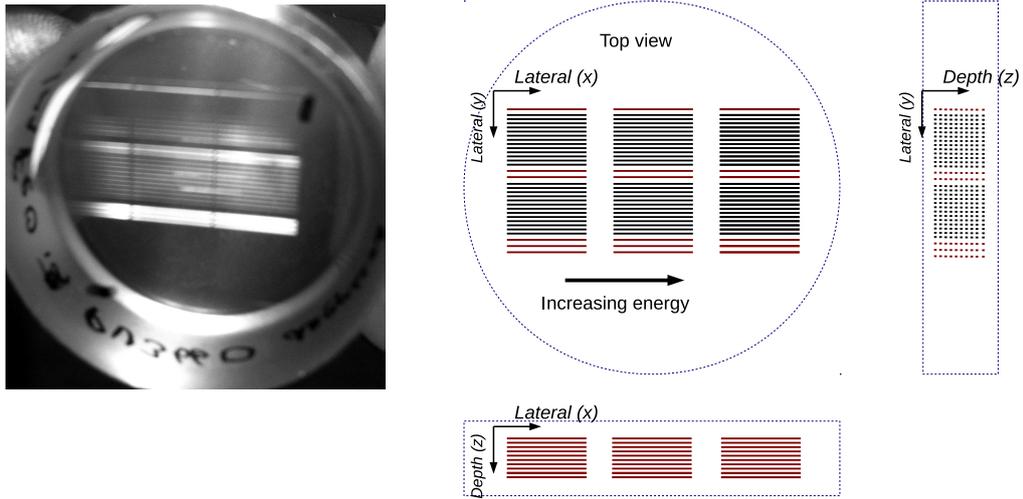


Figure 5.15: Photograph (left) and schematic layout (right) of parameter space test inscribed in FeO_2 particle embedded resin phantom.

Results and discussion

Figure 5.16 shows the lateral variation of structures with increasing energy, inscribed by *Low-Rep 1026 nm* system, for a single depth of $100 \mu m$ and speed of $1 mm/s$. Figure 5.17 shows the lateral variation of the structures with increasing depth for seven different inscription energies. Measurements were obtained from the optical microscope images taken with a $20\times$ objective. A more accurate structural size may be estimated from the phase measurement.

The results show that the size of the modified region increases with increasing laser energy. Similar results were observed for borosilicate as well as for structures appearing slightly more uniform than those observed in the fused silica. Results also indicate that the size of the structure decrease with increasing depth. For lower energies, variation of the size with depth is relatively small compared to those structures written at higher energies. All of the structures inscribed by *Low-Rep 1026 nm* are either damaged or micro voids.

CHAPTER 5. PARAMETER SPACE TEST

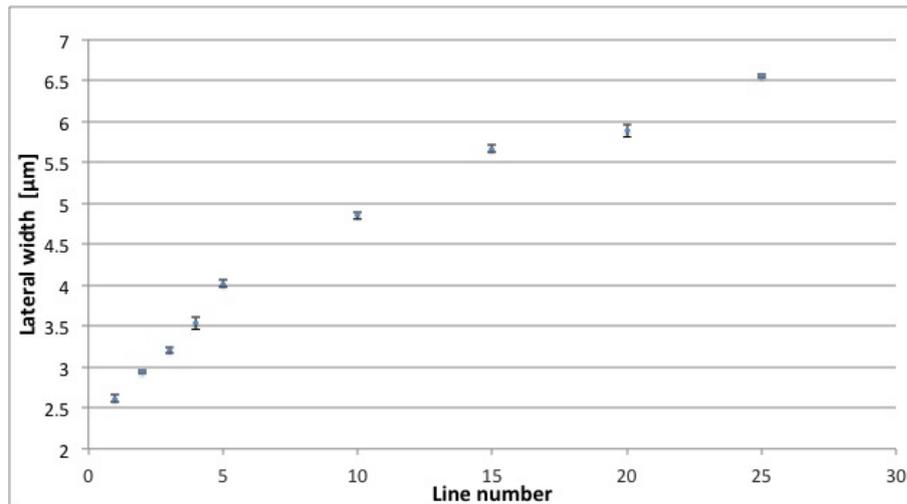


Figure 5.16: The lateral size of the structures inscribed with varying energy for constant speed and depth. Images were obtained using an optical microscope with $20\times$ objective [145]. Energy 1-15 indicates the index of the lines pairs, each pairs were inscribed with different energies. Results given here represent averages over the width from both lines in the pair.

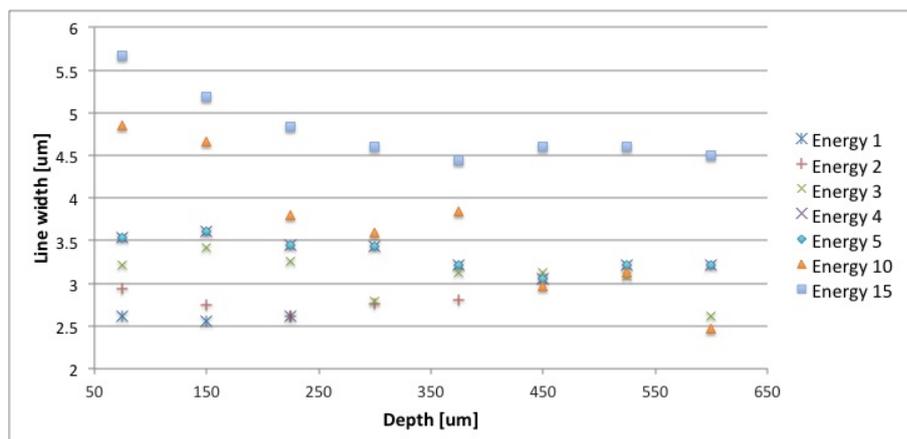


Figure 5.17: The lateral size of the structures inscribed versus depth for constant speed and energy. $E_1 \rightarrow E_{15}$ represent increasing inscription energy.

Figure 5.18 shows the cross sectional profile of the structures (B-scan) of the first five line pairs inscribed with the lowest laser energy in fused silica at 1 mm/s translation speed for varying energy and depths. Figure 5.19 shows the FWHM axial and lateral width of the structures inscribed at 75 μm below the surface. Lines within the pairs are nearly identical, showing the reproducibility of the structures. Furthermore, by calculating the centroid of each point in each frame, linearity of the lines can be estimated. It can be seen that structures

CHAPTER 5. PARAMETER SPACE TEST

inscribed at lower energies look more symmetrical than those at higher energies (Figure 5.18 and 5.19).

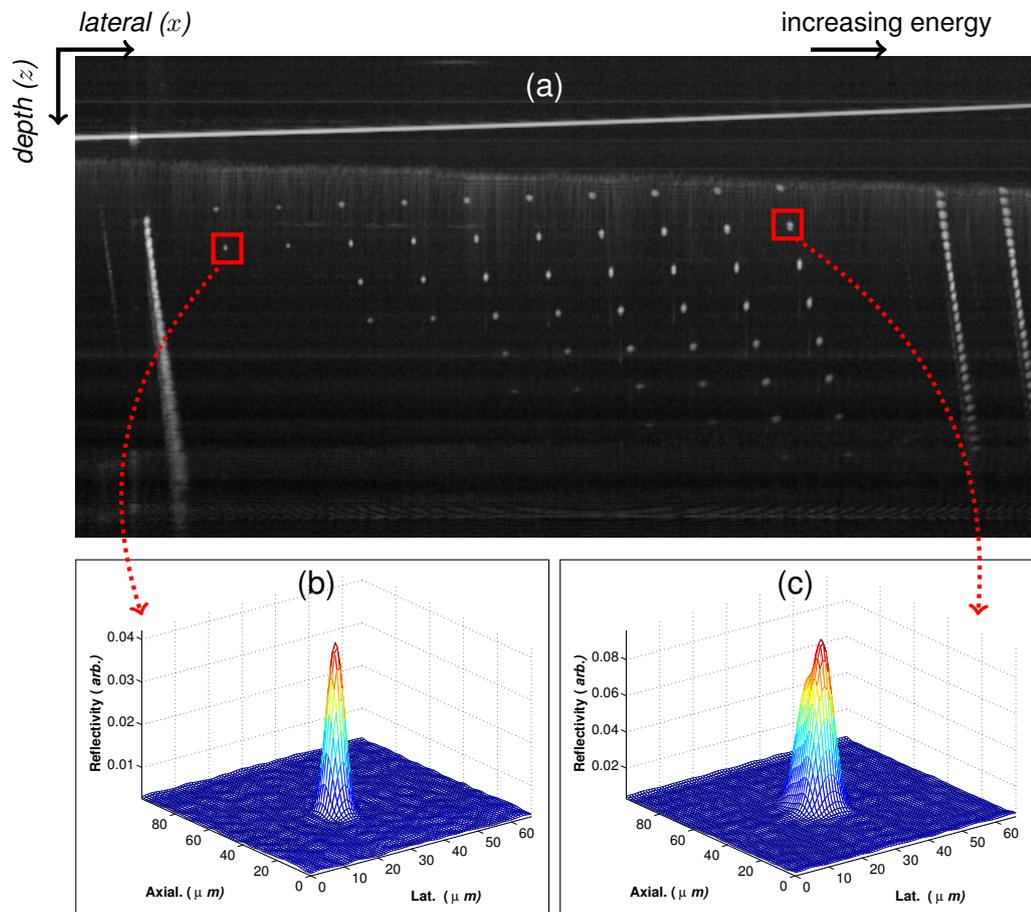


Figure 5.18: OCT results of the first 10 lines from a PST phantom fabricated by Low-Rep 1026 nm . The B-scan of the first 5 pairs of the lines inscribed with lowest energy (a). The mesh plots (b,c) show the zoomed view of the regions marked by the two red rectangles in the B-scan. Mesh plots (b and c) show the structural variation with increasing energy.

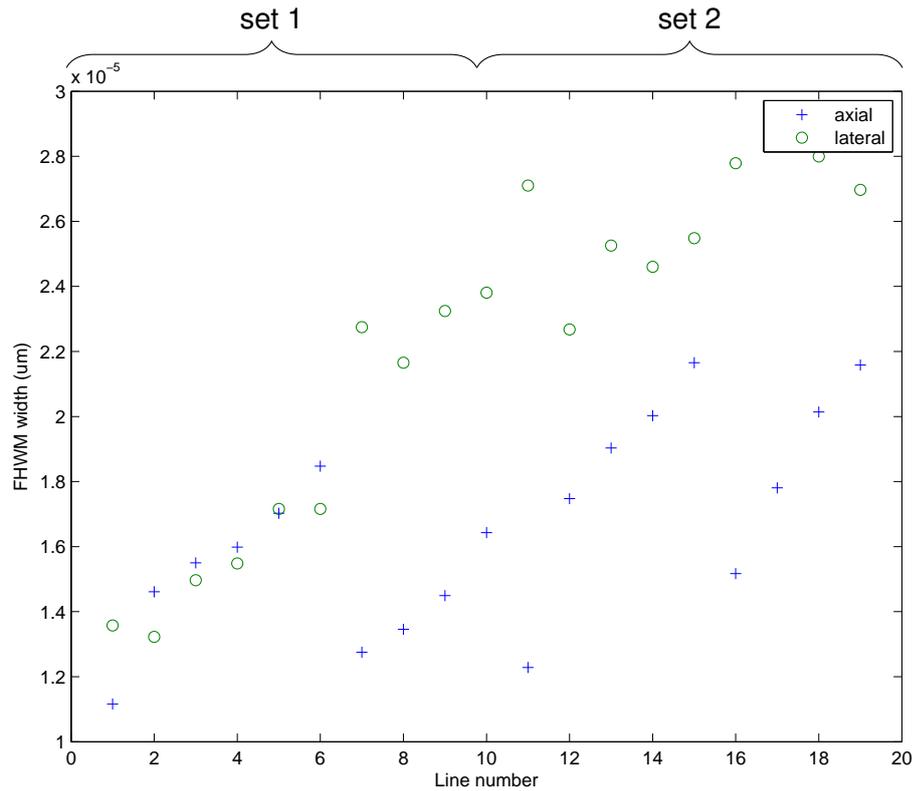


Figure 5.19: The FWHM axial and lateral width of the first layer, obtained from OCT B-scan data. The set 1 is the group of lines shown in 5.18 and set 2 is the next group on the phantom.

5.2 Lower Repetition 800 nm system

Figure 5.20 illustrates the schematics of the parameter space test phantoms inscribed using the *Low-Rep 800 nm* laser system. A series of lines were inscribed at varying speed, energy and depth. Structures were inscribed at 21 different energy values for 11 different speed values for and 10 different depths. The energy was varied along the y-axis from 16.89 to 80 mJ and the speed was varied along the x-axis from 0.1 mm/s to 2.0 mm/s. Structures were inscribed at 10 different depths between 100 μm and 775 μm with a layer separation of 75 μm. The values chosen were selected because of the results of earlier trials. Samples were inscribed in fused silica and PMMA. This section presents characterisation of the parameter space test of these samples using an SS-OCT system and an OM.

CHAPTER 5. PARAMETER SPACE TEST

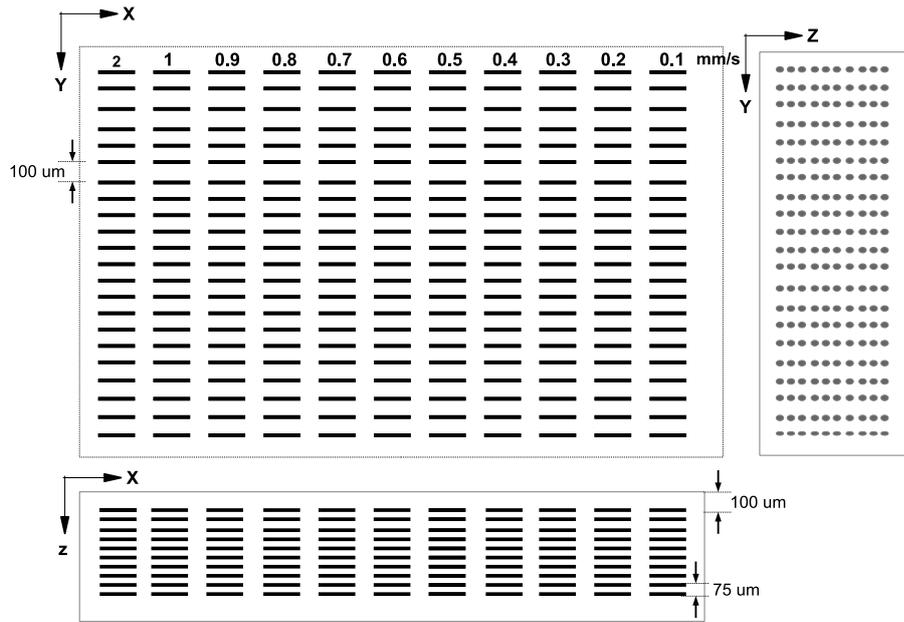


Figure 5.20: Layout of the PST phantom for Low-Rep 800 nm system.

The size of the artefacts produced by the *Low-Rep 800nm* system were limited by the range of the z-axis to 4 mm. In addition, z-axis is a mechanical stage with low precision. Inscription of the structures along the z-axis produce poor results. This makes it inconvenient to produce large calibration phantoms on this system with current set-up. The experimental set-up should be re-configured to produce larger phantoms, more suitable for OCT calibration and characterisation.

Table 5.4: Inscription parameters used for the PMMA and fused silica PST artefacts employing the Low-Rep 800 nm femtosecond system. The energy range chosen here was the results of preliminary investigation conducted with a large range of values.

	Materials	
	PMMA	Fused Silica
Below surface(μm)	75	75
Line spacing(μm)	100	100
Layer spacing(μm)	75	75
Layer offset(μm)	25	25
Line length(μm)	500	500
Energy (nJ)	30 - 178	28 - 136
Depth (μm)	0 - 825	
Speed (mm/s)	0.1 - 2.0	

Nevertheless, the parameter-space was studied by inscribing structures in

PMMA, fused silica, and borosilicate, to study the quality of the inscribed structures as function of energy, speed and depth. The detailed properties of the inscribed artefacts are given in the Table 5.4. The energy, speed and depth were varied along z, x and y direction respectively as shown in the Figure 5.20.

5.2.1 Fused silica glass

Figure 5.21 shows the microscope images of the structures inscribed by *Low-Rep 800 nm* laser system. Top row shows structures inscribed at a constant energy of 28.7 nJ , at a depth of $75 \text{ }\mu\text{m}$ below the surface with a varying speed between $1 - 0.04 \text{ mm/s}$. The middle row shows structures inscribed with varying energy $28.7 - 136 \text{ nJ}$ at constant depth of $75 \text{ }\mu\text{m}$ and speed of 0.5 mm/s . Bottom row shows the structures inscribed at varying depth between $75 - 375 \text{ }\mu\text{m}$ below the surface with an energy of and a speed of .

The images in Figure 5.21, clearly show that the produced pattern changes from, a pearl chain like periodic structure to a near smooth line and then to a damage structure at very lower speed. These structures inscribed by translating the sample at 1 mm/s produced a pearl chain like structure, periodic and smooth. Microscope images suggest that the structures inscribed with speeds between $1 - 0.6 \text{ mm/s}$ are useful for OCT-phantom.

Structures inscribed at higher speeds produce a lower contrast compared to those inscribed at lower speeds, for the same energy and depth. Those made with a lower speed are clearly visible as a line in the microscope images, whereas those formed using a higher speed, give a pearl chain like structure. These present lower contrast and thus are more difficult to see.

Figure 5.22 shows the lateral width of the inscribed structures as function of varying energy, speed and depth. These values were obtained from microscopy images.

⁹Depth of the inscribed lines varies slightly due to the miss alignment in the stage.

CHAPTER 5. PARAMETER SPACE TEST

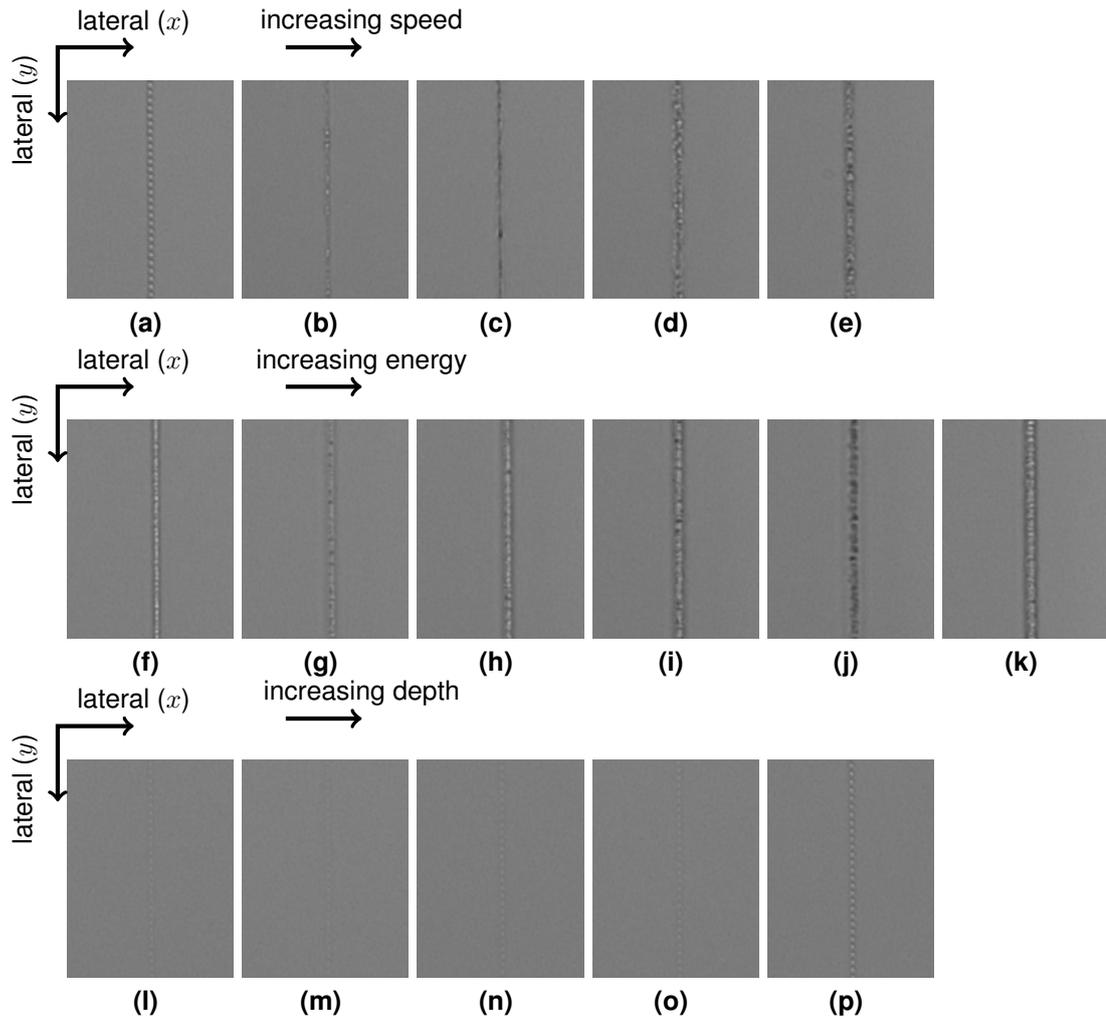


Figure 5.21: Microscope images of the structures inscribed by Low-Rep 800 nm system. Varying speed at constant energy and depth (first row), varying energy at constant depth and speed (middle row), varying depth with constant speed and energy (bottom row).

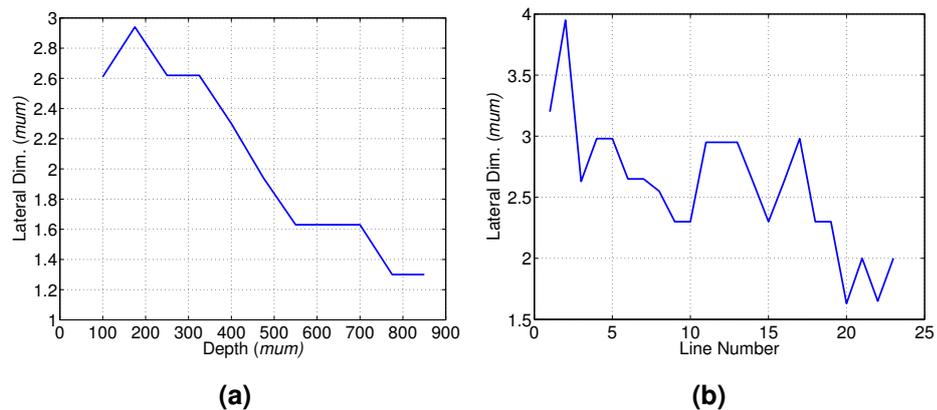


Figure 5.22: Lateral width, obtained from the optical microscope, for a) varying depth at constant energy and speed, b) varying energy at constant depth.

5.2.2 PMMA

This section presents the PST phantoms inscribed by *Low-Rep 800 nm* laser system within the bulk of a PMMA sample. Figures 5.23 show microscope images of a PMMA sample taken by an optical microscope using a $20\times$ objective. First row (a-e) shows structures inscribed at a constant energy of 35.6 nJ , at a constant depth of $75 \text{ }\mu\text{m}$ and at varying speeds of $1 - 0.1 \text{ mm/s}$. Second row (f-k) shows the structures inscribed at constant depth of $75 \text{ }\mu\text{m}$, constant speed of 0.02 mm/s and for varying energy $35.6 - 178 \text{ nJ}$. Third and fourth rows (l-q and r-w) show structures inscribed at varying depth at constant energy and speed. Third row is one inscribed with 35.6 nJ , 0.02 mm/s and the forth row with 178 nJ , 1 mm/s .

Figure 5.24 shows images of lines inscribed at low energy and low speed and for high energy. Low energy produce lines with smaller line width, but the lower speed causes damage in the material. When using a high energy, low speed case causes collateral damage, slow speed also contribute to this.

For the structures inscribed at a higher energy and a higher speed, the fluence is smaller, hence this create less damage, and periodic structures are produced. However, the depth of the inscribed structures are limited to $\approx 200 \text{ }\mu\text{m}$ (i.e. second row).

CHAPTER 5. PARAMETER SPACE TEST

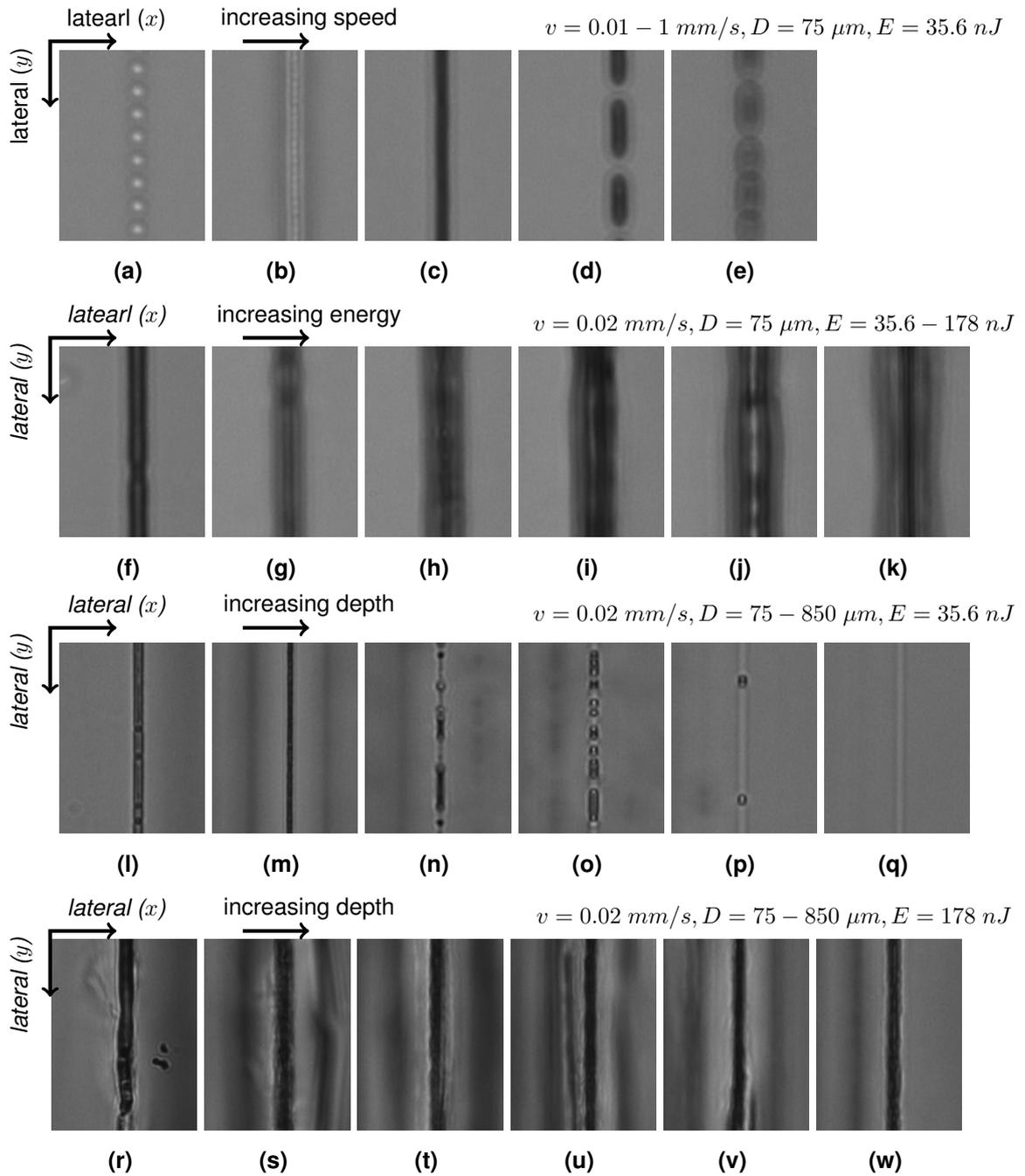


Figure 5.23: Microscope images of the structures of the structures inscribed by Low-Rep 800 nm system. (a-e) inscribed at a constant energy and depth with at varying speed. (f-k) varying energy at constant depth and speed. (i-q) and (r-w) inscribed at varying depth with constant energy (E) and speed (v).

CHAPTER 5. PARAMETER SPACE TEST

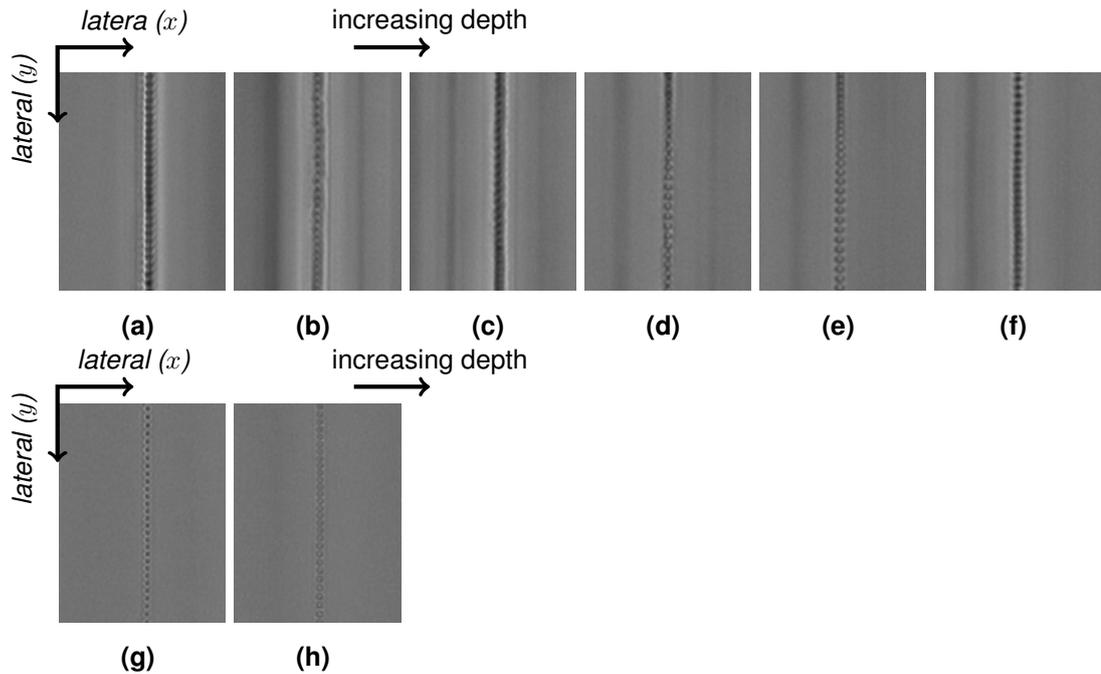


Figure 5.24: Microscope images of the structures inscribed in PMMA with high speed value of $v = 1 \text{ mm/s}$ and a high energy value of $E = 178 \text{ nJ}$ (a-f) and with a low energy value of $E = 35.6 \text{ nJ}$ (g-h). Zoomed in region (i) shows detailed pearl chain like structure with possible position of B-scans(while vertical lines).

Figure 5.25 shows the B-scan of the parameter space test inscribed inside a PMMA by the *Low-Rep 800 nm* system. The row of points near the surface are due to image artefacts. The results were obtained by scanning across the structures inscribed with a translation speed at 0.6 mm/s . The inscription energy and depth of the structures were also varied. The energy decreases from left to right. The depth of layers changes slightly, due to a misalignment in the stage. The missing points in the ninth column are due to a temporary laser malfunction.

CHAPTER 5. PARAMETER SPACE TEST

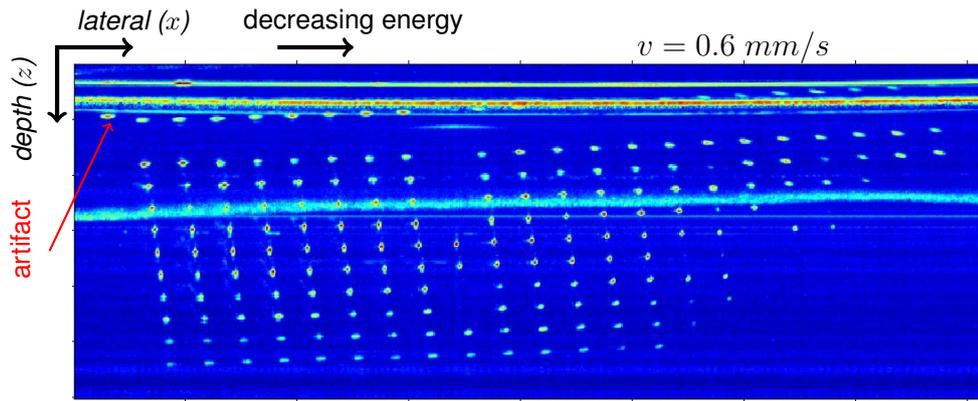
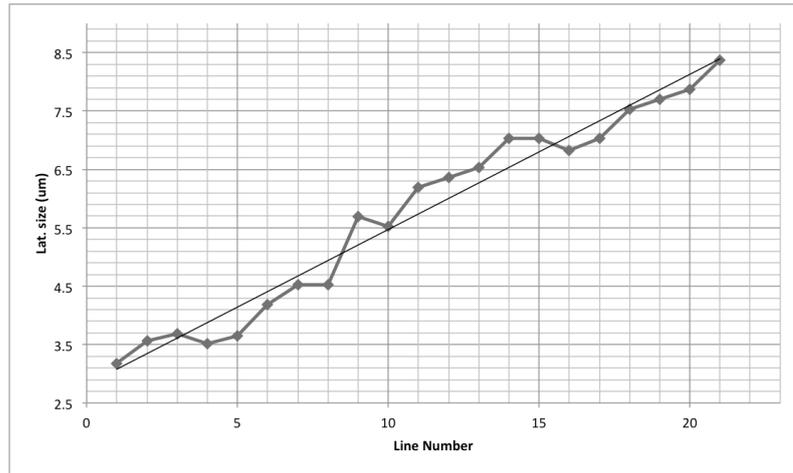


Figure 5.25: B-scan of the PMMA PST sample, where structures were inscribed with varying energy, depth for constant scanning speed. Structures visible near the surface are image artefacts.

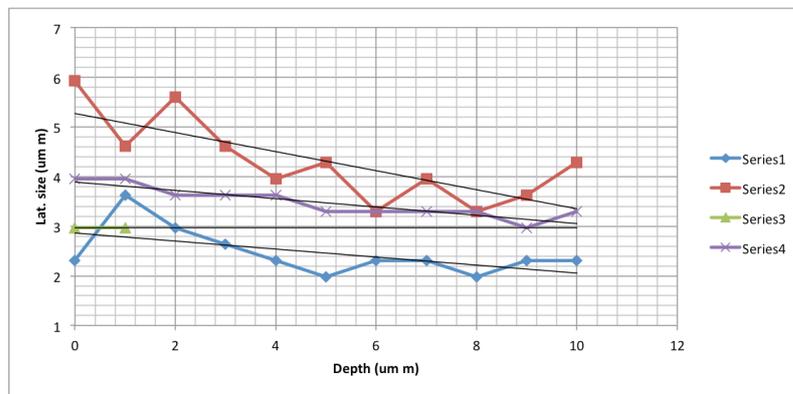
Figure 5.26 shows the lateral size of the structures inscribed in the PMMA by the *Low-Rep 800 nm* system. Figure 5.26 shows the variation of the lateral size with inscription energy (line number), depth and speed. These measurements were obtained using an OM.

Figure 5.27 shows the FWHM axial and lateral width of the inscribed structures versus the inscription laser energy. These data were obtained from the OCT B-scan, from the top layer (i.e. at $75 \mu\text{m}$ below surface).

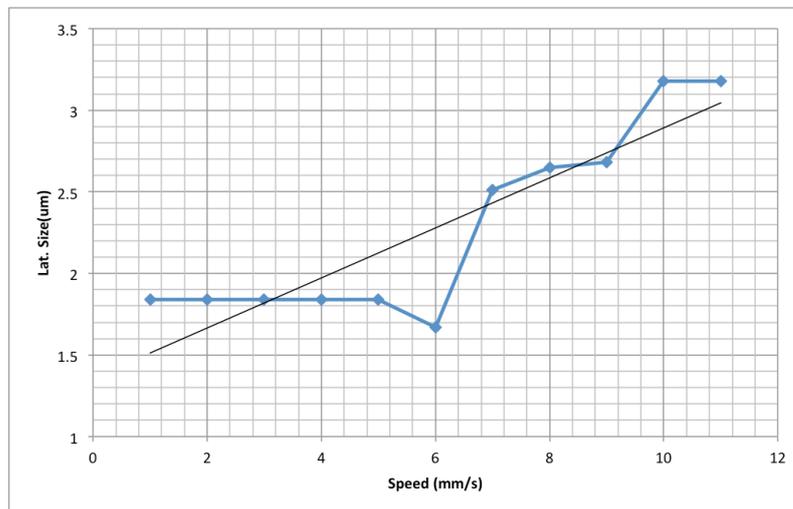
CHAPTER 5. PARAMETER SPACE TEST



(a)



(b)



(c)

Figure 5.26: Microscope measurement of the structures inscribed by the Low-Rep 800 nm inside PMMA. Lateral size variation with a) varying inscription energy, b) varying depth and c) speed.

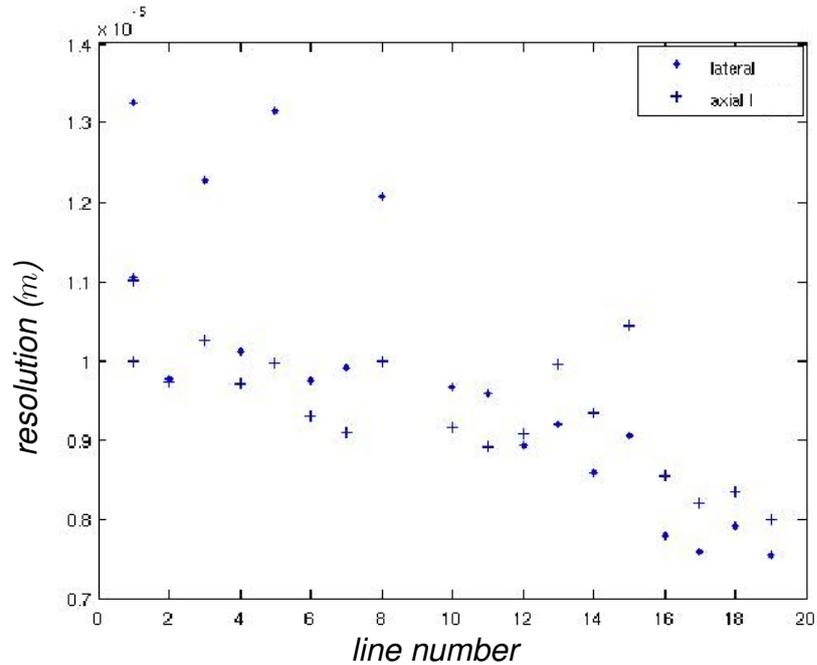


Figure 5.27: FWHM axial and lateral width of the structures, inscribed in PMMA, measured from OCT B-scans.

Individual points are produced at a higher speed ($v \geq 1 \text{ mm/s}$). The contrast (visibility) of the points increases with increasing inscription energy. The size of the structure remains almost constant. However there may be small changes in the size with increasing energy that was unmeasurable using the optical microscope. Studying the cross sectional morphology of the structures using techniques such as SEM and measuring the refractive index using a more accurate technique such as near field interferometry would be interesting.

5.2.3 Discussion

The structure inscribed in PMMA produced a smooth line around 0.7 mm/s , whereas in other materials it produced a pearl chain like periodic structure. This could be attributed to the softness and low melting point of the PMMA. At higher speeds the structures are more periodic, but at lower speed values, for the same energy, damage is observed rather than pure index change. Structures inscribed at higher speed shows undulation, in the subsequent OCT B-scans images taken

along the inscribed line. This could be attributed to the pearl chain like structure produced at high translation speed. However, the PMMA structures present a smaller dimension and are more consistent, hence they are more suitable for an OCT-phantom. The effect of undulation can be minimised by taking average of multiple frames. The phase induced by Low-Rep800 changes from negative to positive with increasing speed due to different processes involved (colour centres and densification [146]).

5.3 High repetition 800 nm system

Figures 5.28 and 5.29 show the microscopy and OCT images respectively of the lines inscribed, inside a doped BK7 glass, at varying energy by a *High-Rep 800 nm* system at constant speed and single depth. When this sample was cleaved and the cross sectional profiles analysed using a microscope, it could be seen that a circular modification was present. The reconstructed refractive index profile showed a wave guide like feature (i.e. high refractive index core surrounded by low refractive index cladding). Details of this work can be found elsewhere [141]¹⁰.

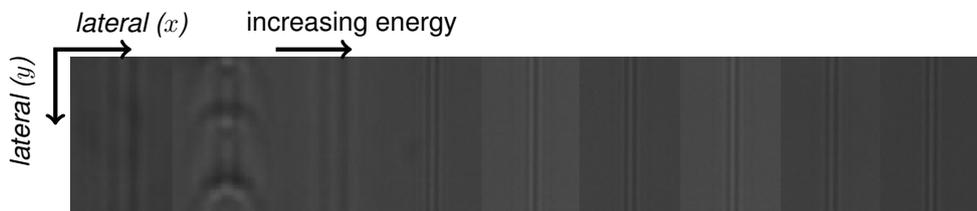


Figure 5.28: Microscope images of the wave-guides inscribed in doped-BK7 glass by a *High-Rep 800 nm* system.

¹⁰Samples were produced by M. Dubov and the results were published in [140, 141]. Refractive index measurements were obtained by myself, the computations of refractive index were carried out in Matlab based on the Mathcad work sheet given by M. Dubov.

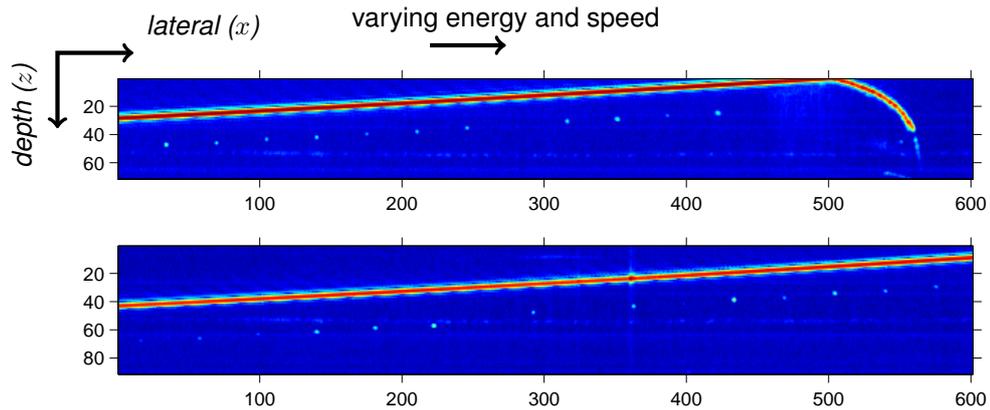


Figure 5.29: B-scan OCT images show sample of the waveguides inscribed with *High-Rep 800 nm fs* system.

5.4 Discussion

Using a home built OCT system, described in section §10.1, it is possible to check the B-scan and *en-face* images in real time. The structures inscribed in the fused silica by the *Low-Rep 1026 nm* system appear as bright spots and lines in B-scan and *en-face* images respectively. The structures inscribed in borosilicate appear as either black lines (*Low-Rep 1026 nm*) or very weak lines (*High-Rep 800 nm*) in the *en-face* image, depending on the system used for fabrication. The structures inscribed in borosilicate, by *Low-Rep 1026 nm* and *High-Rep 800 nm* are not clearly visible in OCT B-scan.

In the lower repetition system, the time between the successive pulses is longer than the time taken for the deposited energy to diffuse out of the focal volume ($1 \mu\text{m}$). Therefore material modification is caused by interaction of individual pulse with the material and amplifies the structural changes, leading to higher contrast structures [39].

In higher repetition systems, the time between pulses is shorter than the time required for absorbed energy to dissipate out of the focal region. Thus, heat accumulation occurs, which causes thermal melting of the material surrounding the focal regions, which produces symmetrical and smoother structures. The

modified region has a spherical shape and its size depends on the accumulated fluence and on the repetition rate of the laser [39].

Results show two types of material modification, refractive index change and void (or micro-cracks). Plasma formed by non-linear excitation at the focal regions expands out from the focal volume into the surrounding matrix. The high density of the plasma formed at higher energies causes an explosive expansion that leaves a less dense centre surrounded by a much denser halo [69].

For all materials and in all systems, translating the sample transversely to the beam waist produced structures. This creates linear damages inside the material around the focal regions. Structures inscribed with energies greater than material modification thresholds (i.e. modified regions) are observable by optical microscope (and naked eye). When the structures were inscribed within the substrate sample, the surfaces of the samples appeared unaltered.

The size of the inscribed structures in fused silica, by both lower repetition systems, stays nearly constant for all speed values at constant energy and at constant depth. The size of the structures in PMMA, however, increase with increasing speed. In PMMA it is believed that absorption of multiple pulses increases the temperature of the localised region significantly. This causes decomposition of the polymer substrate at the focal volume. In fused silica it is believed that the observed refractive index change due to material density or chemical change leads to these observed structures.

5.5 Conclusion

In conclusion, this work compares the micro structures fabricated using three different fs lasers and evaluates their suitability as OCT artefacts. The properties of the fs inscribed structures from three different systems were analysed. Structures fabricated within the fused silica, PMMA and doped-BK7 are observable by OCT, whereas those inscribed in borosilicate are not visible in MD-OCT system. However, structures in borosilicate were visible in the *en-face* images obtained using

the RT-OCT system. This could mean that the sensitivity of the RT-OCT system is better.

The results show that fs inscribed structures are highly reproducible at certain fluencies. The densification structures produced by the low repetition rate systems in fused silica and PMMA are better visible in OCT measurements than the wave guide like structures produced by the high repetition system. Structures produced by the high repetition system are more symmetrical and homogeneous than those produced by low repetition systems. The size of elongation increases with increasing energy; however the use of a cylindrical lens or slit can reduce this effect. Structures written with the *Low-Rep 800 nm* system in PMMA tend to be more symmetrical than those in fused silica, perhaps because of the relaxation of the stress in the PMMA.

In conclusion, fabrication of fs laser inscribed PMMA artefacts is demonstrated. It is shown that this approach has the potential to inscribe 3D PSF artefacts with sub-micrometer accuracy. Working in PMMA is lower cost than in silica and the structures written appear to be smoother. Additionally PMMA can be easily moulded to emulate shapes similar to those usually found in biological samples.

Although the induced refractive index changes of the structures inscribed in borosilicate using the *High-Rep 800 nm* system were much larger than those in doped-BK7, they were not visible with the OCT system, while clearly visible using an OM. This is probably due to sample quality or material differences but this is still under investigation.

Both *Low-Rep 800 nm* and *Low-Rep 1026 nm* laser system produce $1 \mu m$ structures limited by the focusing conditions. Inscription depths of the smallest structure inscribed by both low rep systems are limited to few hundred $\mu m s$. No structural modifications were observed at deeper depths, even for lower inscription energies. Also, some of the fine structures visible in microscope images were not resolvable under OCT. This could indicate the sensitivity limit of the OCT system. This limit could be due to the focusing conditions of the objective used, as

CHAPTER 5. PARAMETER SPACE TEST

well as the need to choose the central wavelength. The contrast of the inscribed structure increases with increasing energy.

In conclusion, properties of these structures can be well characterized. The dimensions and positions of, as well as induced refractive index in the fs inscribed micro structures can be well controlled by controlling the inscription parameters. Therefore it is realistic to produce standardization artefacts in this way with designs suitable for calibrating and characterizing OCT and other imaging systems.

Chapter 6

Development of characterisation phantoms

This section gives detailed descriptions of the designs, fabrication methods and characterisation of calibrated phantoms to be used in OCT characterisation. The OCT-phantoms presented here were designed to qualitatively and quantitatively evaluate important parameters control requirement of OCT systems namely resolution, sensitivity, distortion and scan linearity.

The inscribed structures were characterized using an optical microscope (Axioscope 2 MOT pulse, Zeiss) equipped with quantitative phase microscopy and an SS-OCT (Michelson Diagnostic, EX1301) system. The following sections include a description of the design requirement, design of phantoms and some preliminary results. Also, an overview is presented on how to measure the performance indicators using the OCT-phantoms.

6.1 PMMA - Single layer

The *Low-Rep 800 nm* laser system was used to inscribe lines at a marginally varying depth inside PMMA (Clinical grade, Vistacryl CQ non CV) samples of $10 \times 10 \text{ mm}$ size and 5 mm thickness. After characterising the sample, an optimum

set of writing parameters was chosen to write the PSF artefact. The PSF artefact was written at a 31 nJ energy by translating the sample at 0.7 mm/s . A total of 59 lines with a $25 \text{ }\mu\text{m}$ pitch were inscribed between 100 and $112 \text{ }\mu\text{m}$ below the surface, starting from left to right (i.e. equivalent to 1.2 degree tilt). The depth was chosen as a compromise between the quality of the structure and the level of interference due to the strong back reflection from the front sample surface. At deeper levels, the structures become uneven or the sensitivity is not high enough for the OCT system to detect.

Schematics of the design (a), microscope image (b) of eight lines, a B-scan image (c) and *en-face* image (d) of the phantom are shown in Figure 6.1. Microscope image was obtained with a $5\times$ objective. A 3D reconstruction of the phantom is presented in Figure 6.2. Lateral width of the inscribed lines, measured from microscope image, is $2.61 \text{ }\mu\text{m}$. The inscribed line widths are smaller than $3 \text{ }\mu\text{m}$, thus the effect of the width on the PSF could be neglected.

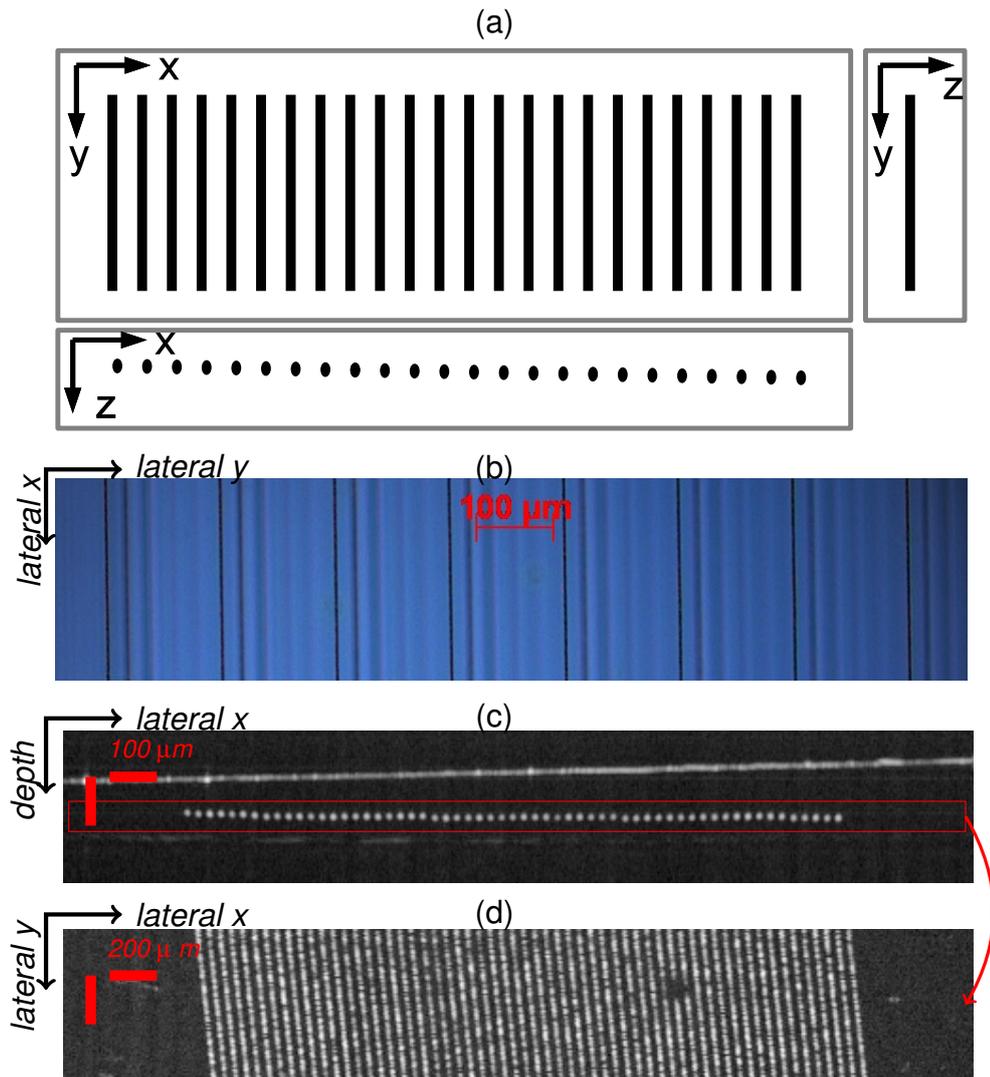


Figure 6.1: The single layer PMMA phantom. A schematic of the phantom (a), microscope image of 8 lines (b), B-scan (c) and an en-face (c) images. Lines were inscribed at a depth between $100 - 112 \mu\text{m}$ below the PMMA surface, with a pitch of $25 \mu\text{m}$. Structures were inscribed using the Low-Rep 800 nm with an energy $E = 31 \text{ nJ}$ insert energy value at a speed of 0.7 mm/s .

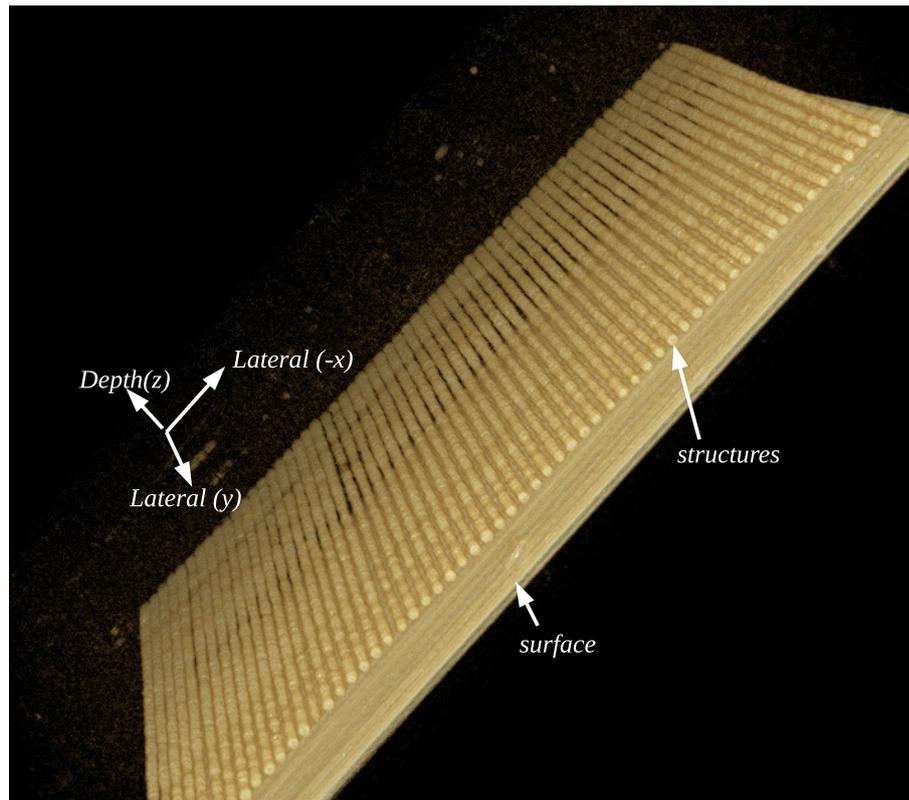


Figure 6.2: Three dimensional rendering of PMMA single layer phantom

All the points in the PSF artefacts are confined to the best operating region of the system. Therefore the effect of axial resolution degradation as a function of depth can be neglected. PSF of an OCT system, with axial and lateral resolution $> 8 \mu m$ could be directly measured using phantom. Single layer avoids interference effects between multiple layers.

6.2 Fused Silica - Multi layer

6.2.1 General design

The work culminated in the design of OCT-phantoms containing four geometrical test patterns. Each pattern was designed with specific test functionality.

All four test patterns were fabricated within the bulk of a $25 \text{ mm} \times 30 \text{ mm} \times 2 \text{ mm}$ fused silica window (Ibsen Photonics, Denmark) using the *Low-Rep 1026 nm* fem-

CHAPTER 6. DEVELOPMENT OF CHARACTERISATION PHANTOMS

to second laser system. Fused silica was chosen due to its excellent optical transmission bandwidth which makes it ideal for use on OCT systems operating at different wavelengths. PST results show that structures inscribed in fused silica, by *Low-Rep 1026 nm* laser system, have higher contrast than Borosilicate¹.

Figure 6.3a shows a schematic of the overall layout of the final phantom design that consists of four test patterns; these are named from left to right: PSF (P), Sensitivity (S), Lateral resolution (L) and Distortion (D). A photograph of this final phantom is shown in Figure 6.3b.

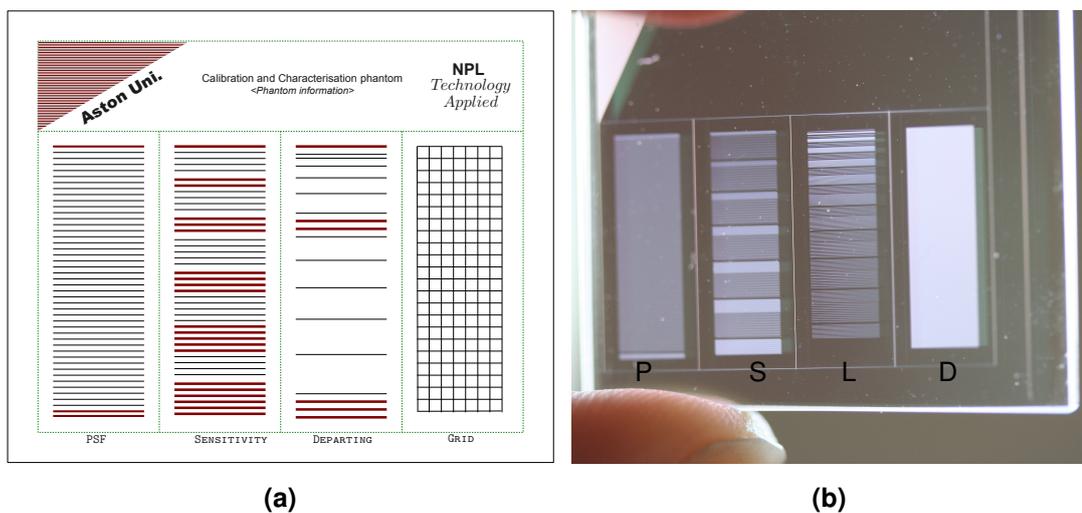


Figure 6.3: Schematics of the overall layout (front view) and the photograph of the OCT calibration and characterisation phantoms. P: PSF; S: sensitivity; L: lateral resolution; D: distortion.

All four test patterns presented have eight layers inscribed at a translation speed of 10 mm/s . The top layer was inscribed $100 \mu\text{m}$ below the sample surface and the subsequent layers were inscribed with $75 \mu\text{m}$ spacing in between them. The PSF, lateral resolution and distortion structures were inscribed with an average pulse energy of 860 nJ ($0.5E_T$). The Sensitivity pattern contains 30 pairs of lines that were written with a pulse energy between 380 nJ to 1.3 mJ per pulse. In the PSF and Sensitivity patterns, the layers were horizontally offset by $25 \mu\text{m}$ in order to reduce multiple back scattering occurring during OCT measurements.

¹We have only investigated fused silica and borosilicate on this system. It would be interesting to study a parameter space test on other glasses and polymers (i.e. PMMA).

In the figures presented in this section, the letters $d_1, d_2, d_3, d_4, d_5, d_6$ and d_7 respectively denote the spacing between the index line and test line, the spacing between the test lines (pitch), spacing between index lines, distance between first layer and surface, layer separation, length of the lines and layer offset.

6.2.2 PSF phantom

The PSF pattern presented here contains a grid of lines inscribed with a constant energy. A total of 110 lines were inscribed at each layer with spacing of $150 \mu m$ between the lines.

Figure 6.4 shows a phantom layout (a), microscope image of first four line pairs (b), OCT B-scan of a section (c) and the en-face image of the second layer, marked by red rectangle in B-scan (c). The microscope measurements give an average lateral width, measured at the first layer, of $4.8 \pm 0.2 \mu m$ and the size of the structure varied slightly with depth. It would be possible to compensate this by changing the inscription energy with depth or using additional optics. Figure 6.5 shows the 3D reconstruction of the distortion phantom using OCT data. Rendering is achieved using Drishti volume rendering software.

CHAPTER 6. DEVELOPMENT OF CHARACTERISATION PHANTOMS

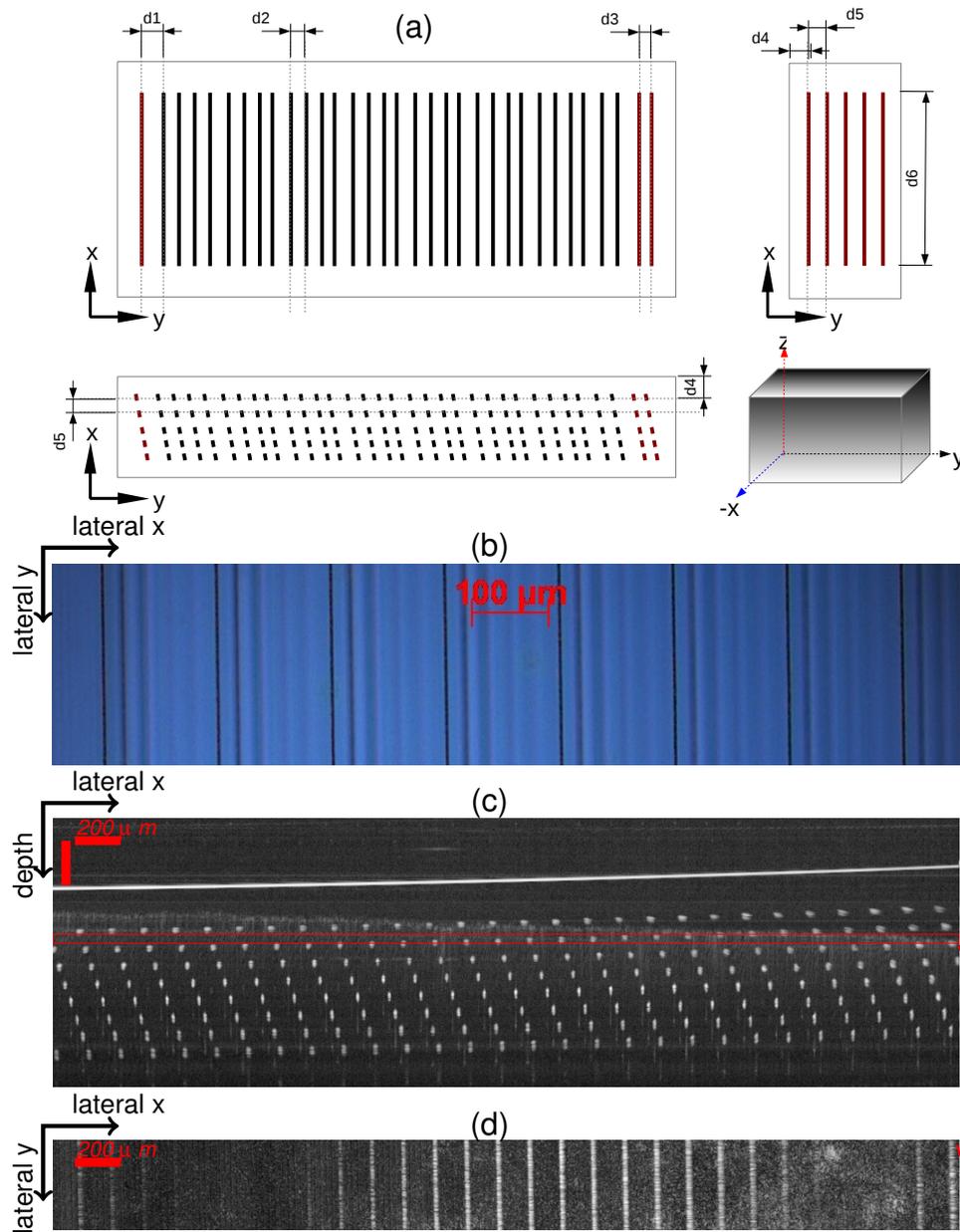


Figure 6.4: PSF phantom. a) Schematics, b) microscope image of the top layer, c) B-scan and d) en-face image of the section shown by the red rectangle in (c).

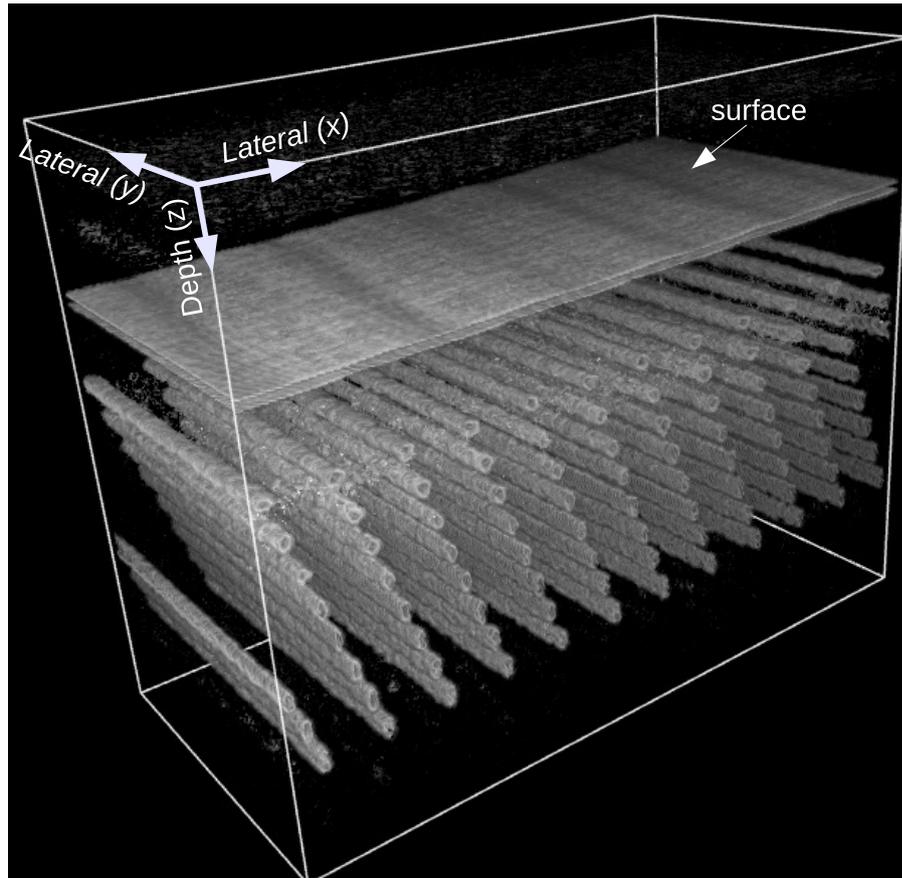


Figure 6.5: Three dimensional rendering of PSF phantom. Rendering is achieved using Drishti volume rendering software.

6.2.3 Sensitivity phantom

The sensitivity phantom was fabricated by inscribing pairs of lines, all with identical inscription parameters. Pairs of lines were inscribed with the same properties in order to minimize the occurrence of false positives caused by spurious phantom signals. The inscription energy between each pair increases from left to right of across the sample. The line pairs were grouped into sets of five pairs and each set is separated by index-lines written with a high laser power. The index-lines help the user locate the location of the sensitivity lines. i.e. the set n refers to the group of lines inscribed between the n and $n+1$ index lines (indicated by red in Fig 6.6a).

Figure 6.6 shows the sensitivity pattern layout (a), microscope image (b) of the

CHAPTER 6. DEVELOPMENT OF CHARACTERISATION PHANTOMS

first four pairs of lines, a B-scan image (c) and the *en-face* image (d) of the first set (first five pairs inscribed with lowest laser energy) of the sensitivity pattern. Each layer consists of six sets with five pairs in each set (i.e. 30 pairs). Spacing between the lines was $150 \mu m$. The line pairs were inscribed with an average pulse energy between $380 - 1300 nJ$ (i.e. 20%-80% of the total pulse energy). Figure 6.7 shows the 3D reconstruction of the grid phantom.

The microscope measurements of the lateral widths are between 1.7 and $6.2 \mu m$.

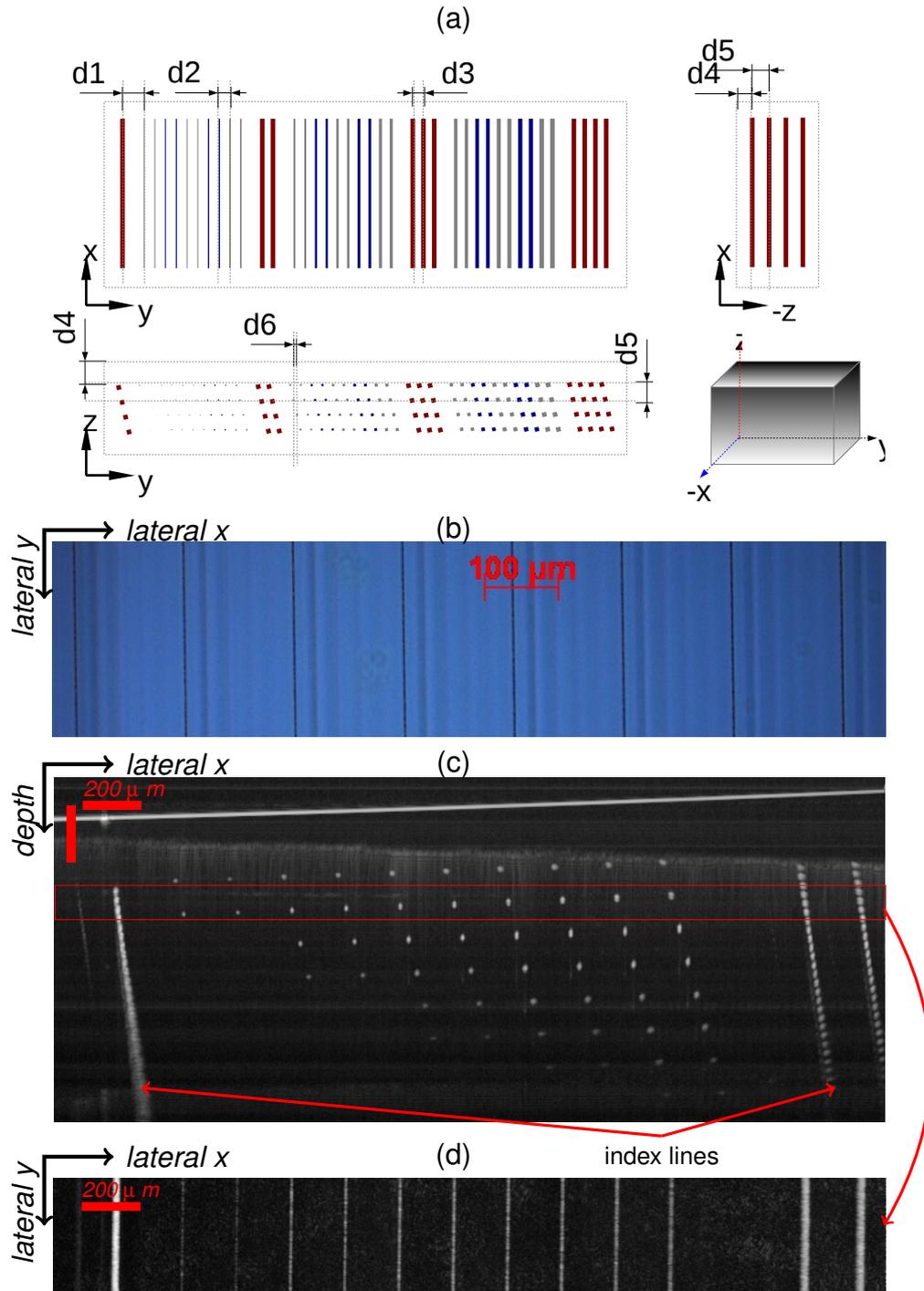


Figure 6.6: Sensitivity phantom. a) Schematics, b) microscope image of the top layer, c) B-scan of a section and d) en-face image shown by the red rectangle in (c).

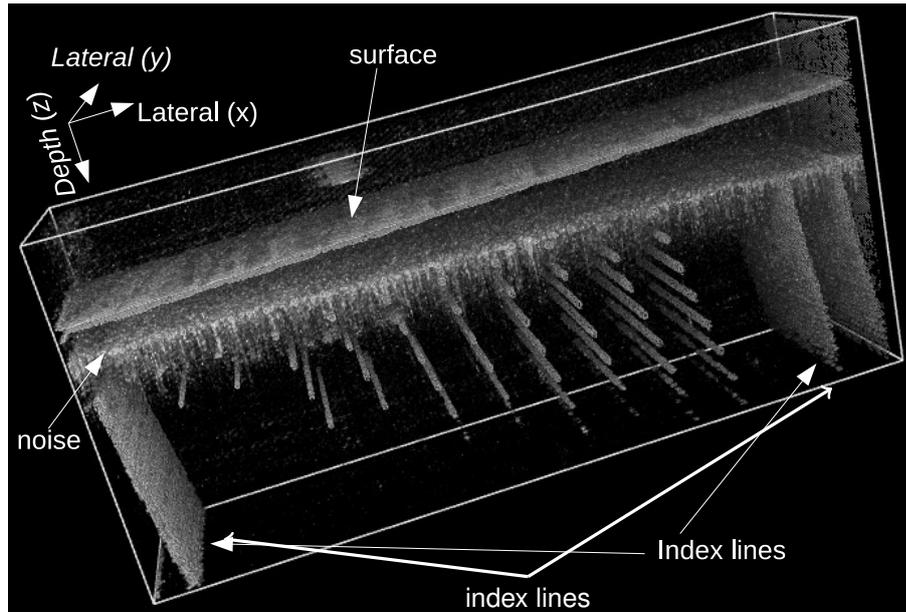


Figure 6.7: Three dimensional rendering of sensitivity phantom. Rendering is achieved using Drishti volume rendering software.

6.2.4 Distortion phantom

Grids of $100 \mu m \times 100 \mu m$ were inscribed at eight different depths. The overall dimensions of the distortion pattern are $525 \mu m$ thickness (i.e. the lowest layer from top surface) $4 mm$ width and $13 mm$ long containing 8 layers, 130 vertical and 40 horizontal lines).

Figure 6.8 shows the design schematics (a), microscope image (b) of the first layer, the B-scan (c) and *en-face* (d) of the distortion pattern. Figure 6.9 shows the 3D reconstruction of the grid phantom.

CHAPTER 6. DEVELOPMENT OF CHARACTERISATION PHANTOMS

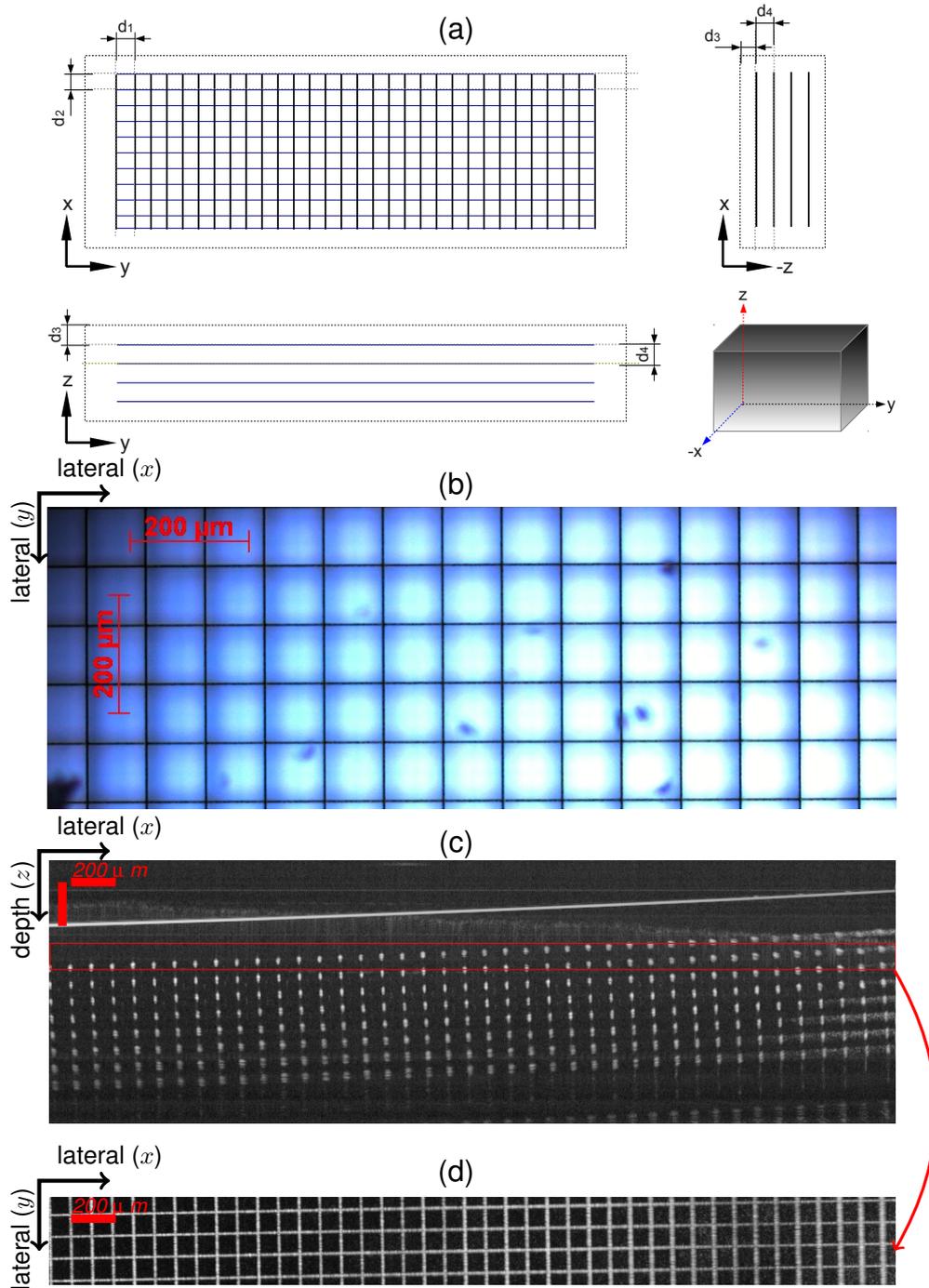


Figure 6.8: Distortion phantom. a) Schematics, b) microscope image of the top layer, c) B-scan of a section and d) en-face image of the section shown by the red rectangle in (b).

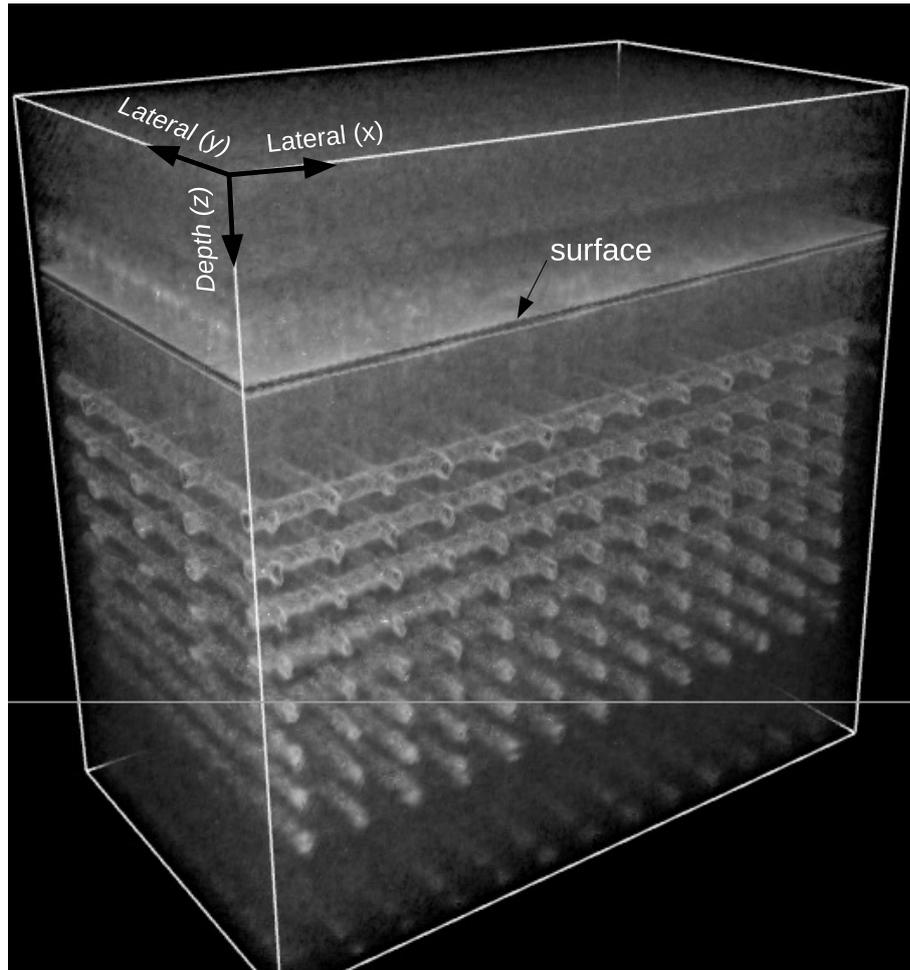


Figure 6.9: Three dimensional rendering of grid phantom, obtained using the Drishti volume rendering software.

6.2.5 Lateral resolution phantom

The lateral resolution pattern was designed so that the spacing between the lines increased linearly - varied from 1 to 110 μm . Each set contains 10 spaces e.g. 1 – 10 μm and 11 – 20 μm , and the size of the line-spacing increases from left to right. Index lines separate each set. As with the sensitivity pattern, set n is separated by index lines n and $n + 1$. The spacing between the lines n and $n + 1$ in a set m can be identified by: $S(m, n) = N(m - 1) + n$, where m and n are set number and line in the set m .

Figure 6.10 shows schematics illustration (a), microscope images of the first

and second set from the top layer (b-c), the B-scan (d), the *en-face* OCT images (e). Microscope images of set 1 (b) and set 2 (c) were obtained using $63\times$ and $20\times$ objective respectively. The OCT images show the first five sets (i.e. line spacing $1 - 50 \mu m$ and the index lines). Figure 6.11 shows the 3D reconstruction of the lateral resolution phantom.

The lateral width of the lines is $4.8 \mu m$. Therefore, a spacing smaller than $4.8 \mu m$ is not possible, otherwise the lines merge together. The size of the inscribed structures depend on the inscription energy and speed. Therefore it is possible to achieve smaller line spacing with a lower energy pulse or a different machining objective [30].

CHAPTER 6. DEVELOPMENT OF CHARACTERISATION PHANTOMS

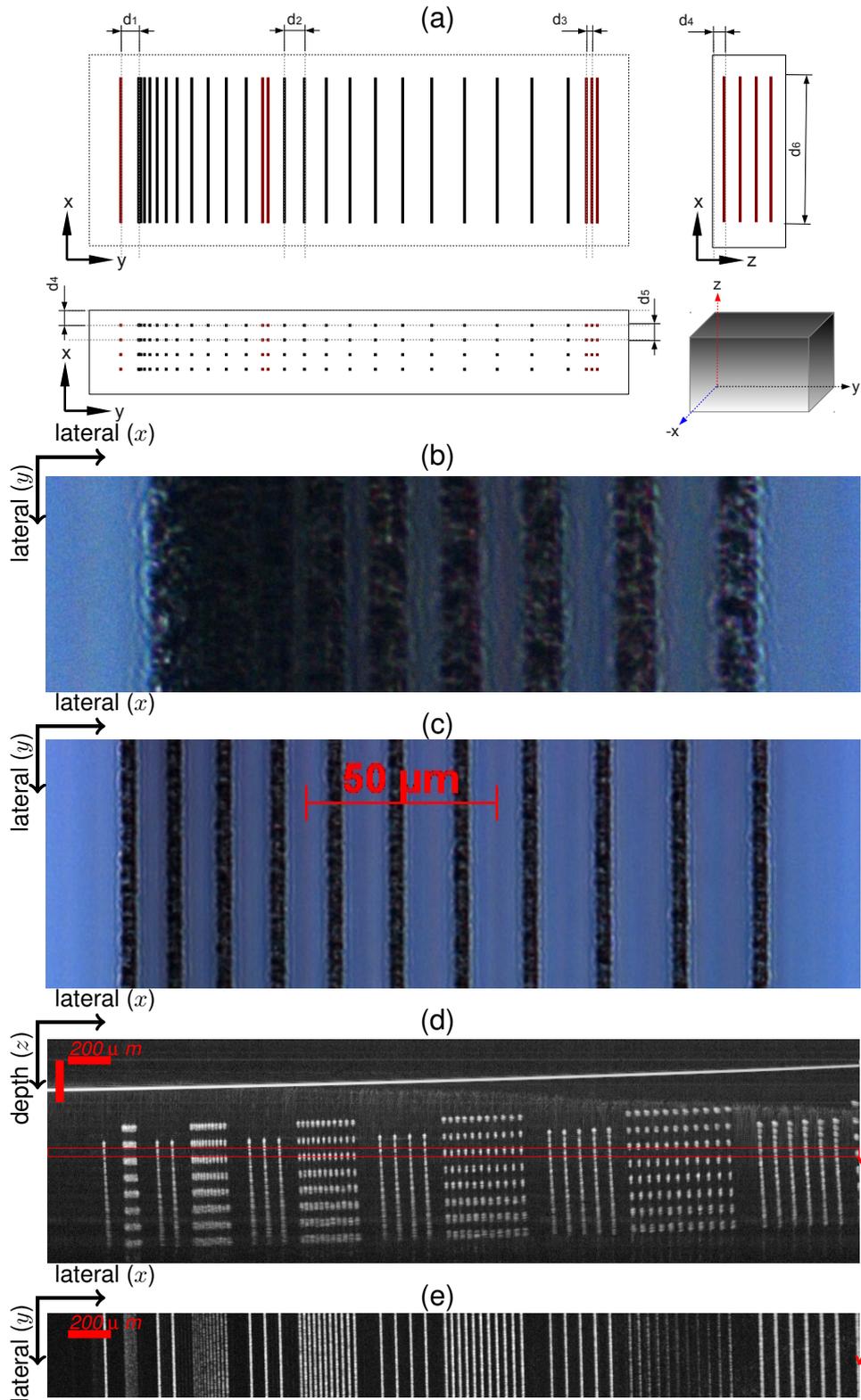


Figure 6.10: Lateral resolution phantom. a) Schematics, b) microscope image of the top layer, c) B-scan of a section and d) en-face image of the section shown by the red rectangle in (b).

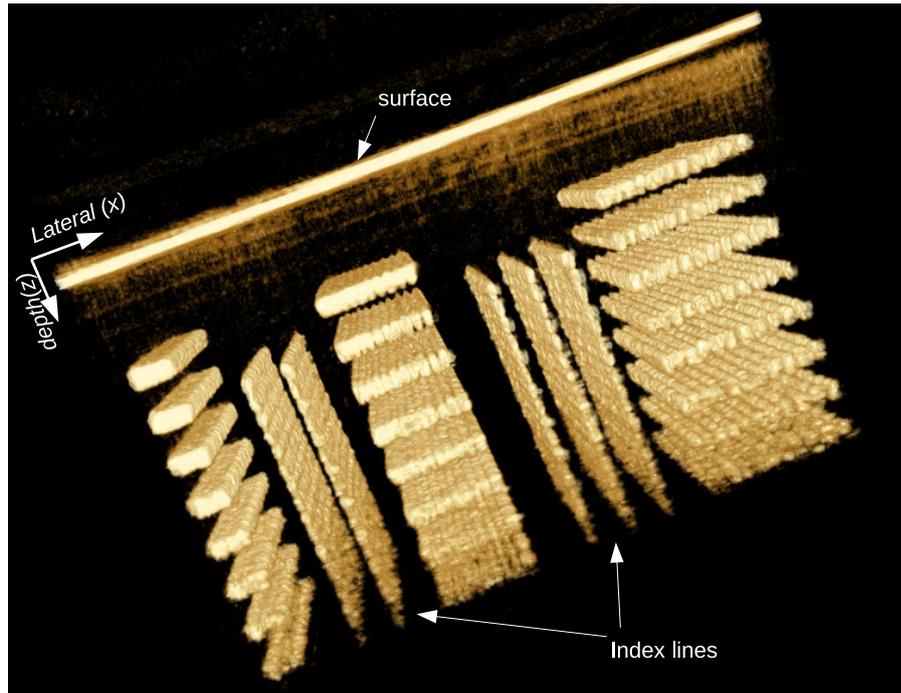


Figure 6.11: Three dimensional rendering of lateral resolution phantom. Three sets of lines (i.e. between $1 - 30 \mu m$) separated by two index lines.

6.2.6 Discussion

The structures (lines) in PSF, Lateral resolution and Distortion phantoms were inscribed with the same energy and speed. The microscope measurements give an average lateral width, measured from the first layer, of $4.8 \pm 0.2 \mu m$ and the size of the structure varied slightly with depth. The OCT measurement gives an average FWHM axial and lateral width of $10.43 \pm 0.40 \mu m$ and of $24.70 \pm 0.70 \mu m$ respectively. This is obtained by fitting a two dimensional Gaussian function to the PSF and distortion patterns. The OCT images also show slight elongation of the inscribed structures in the axial direction as well. It would be possible to compensate for the elongation using additional optics and varying the laser power used for inscription with depth.

The size of the structures inscribed in the current OCT phantom is larger than the size required to directly measure the PSF of ultra-high resolution system, where the resolutions are better than two μm . The fabrication of structures

smaller than the resolution of an OCT system would be possible at lower inscription energies.

The structures in the sensitivity phantoms were inscribed with varying energies. The microscope measurements of the lateral widths are between 1.7 and 6.2 μm . There were no identifiable structures by the microscope below the third layer for inscription energies of 381 nJ (i.e. first pair). This suggests that the fluencies used for these pairs were too low to produce any structural changes below this depth. The number of visible layers increased with increasing energy and the size of the structure decreases with increasing depth. The fabrication of structures smaller than the resolution of the OCT system are possible at lower energy.

6.3 Conclusions for both PMMA and Fused silica

A solid, durable and highly reproducible calibration and characterization phantom has been fabricated using femtosecond laser inscription technique. Four geometrical patterns were inscribed inside a fused silica glass which allows to qualitatively and quantitatively determine several performance matrices of an OCT system such as resolution, distortion, sensitivity, resolution degradation and sensitivity roll-off, imaging depth and scan linearity. The phantom patterns were characterized using optical microscopy and OCT imaging. As the OCT systems evolves, need for calibration and characterisation becomes vital. OCT-phantoms developed during this study will allow both researches and end-users to calibrate their OCT systems. Intra-laboratory and comparison of systems will also be made possible.

Chapter 7

OCT system characterisation using OCT phantoms

7.1 Resolution

The following section describes the development of quantitative methods for experimentally measuring the spatially varying 2D PSF using OCT-phantoms produced as described above. This method can be extended to measure spatially varying 3D PSF volumetric OCT data (3D data).

Previous work has shown that the OCT system used in this study exhibits a non-uniform sensitivity across the lateral scanning and a finite depth of field [31].

7.1.1 Method

Volumetric data were obtained using MD-OCT for both single (PMMA) and multi layered (fused silica) PSF patterns. Data were processed as described in chapter 9. Points in B-scans were identified, extracted, oversampled and smoothed and axial and lateral FWHM of each points were estimated by fitting a 2D Gaussian function. A $7 \times 7 \times 10$ and $21 \times 21 \times 10$ pixels were extracted from single layered and multi layered sample data respectively.

7.1.2 Results and Discussion

PMMA

Figure 7.1 shows the histograms of the measured FWHM of axial and lateral PSFs for all of the measured points. As expected, the pronounced symmetry indicates that all the points lie within the focal and scanning range of the OCT system. Therefore, we can assume that the effects of distortion are negligible here. The histogram peak locations correspond to mean FWHM PSF values of $7.4 \mu\text{m}$ and $7.8 \mu\text{m}$ for the axial and lateral resolution respectively.

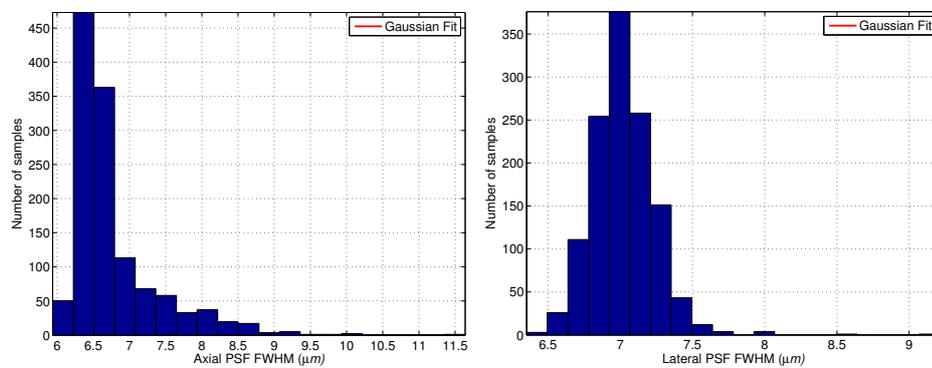


Figure 7.1: Histograms for the FWHM of (a) lateral and (b) axial PSF.

The sample is also tilted in the A-C plane by 6.47^{deg} degrees. In order to avoid strong back reflection from the front surface, the sample was tilted in the xy plane, which cannot be seen in the xy plane. This in effect changes the depth of the lines with varying B-scan, and as such the lines move away from the focal point in that plane. This gives the observed variation of the lateral PSF along the axial direction.

The sample tilt has an additional benefit since it means that the position of the lines varies in each frame relative to the focus. This allows the effect of depth on the resolution to be studied as a function of axial distance. The results in Figure 7.2 show the variation of lateral and axial resolution as a function of depth. The axial resolution (green) is constant with varying depth whereas the lateral resolution (blue) varies significantly. This is in agreement with the previous stud-

ies [31, 114].

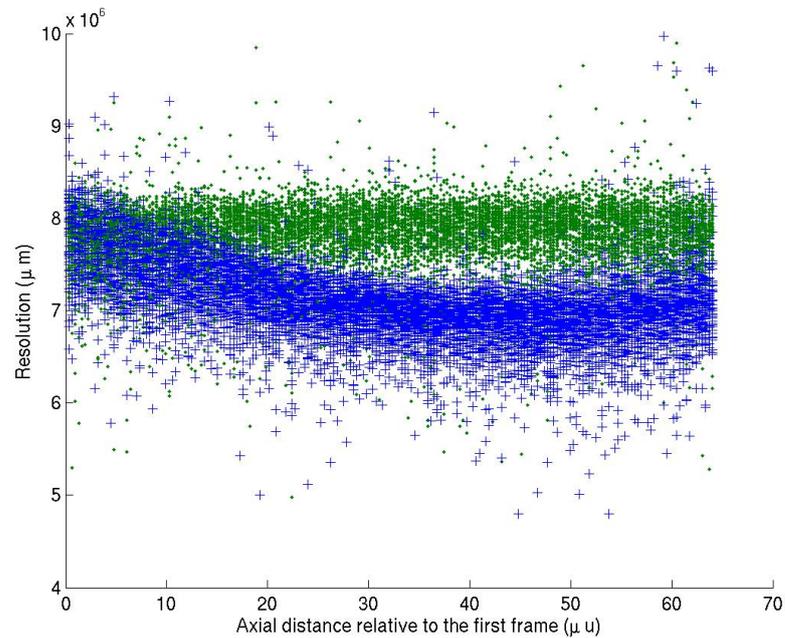


Figure 7.2: Measured FWHM lateral (blue) and axial (green) PSF as a function of axial position.

Fused silica

A $21 \times 21 \times 20$ pixels were extracted around each point and points from closely spaced sequential B-scans were oversampled and smoothed as described in section 3.7. The axial and lateral FWHM widths of each of the points in the B-scan were estimated by fitting a two dimensional Gaussian function (given in equation A.0.1). Figure 7.3 shows the axial and lateral FWHM widths, as function of axial and lateral direction and their distribution.

The asymmetry of the distribution in lateral resolution suggests that the data were obtained over a depth that extends well outside the DOF of the system. In this particular case, the beam was roughly focused at the second row. The focal depth of $250 \mu m$ only covers 3 – 4 rows, hence other points will lie outside the focusing region.

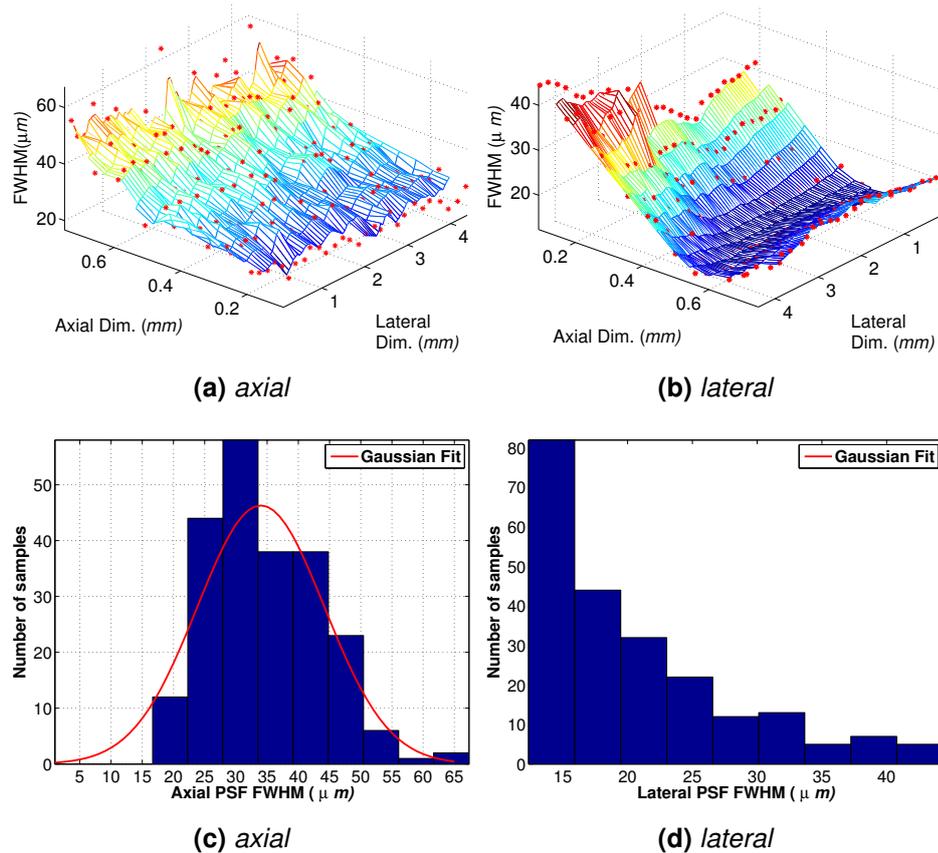


Figure 7.3: The variation of the a) axial and b) lateral PSF as a function of axial lateral position in the B-scan. The results were obtained by oversampling the peaks in each slice of the sub-volume to common coordinate and then fitting a 2D Gaussian.

The statistical distributions presented in the histogram (Fig. 7.3c) show axial symmetry, with only one side of the Gaussian visible due to the fact that the OCT was focused near the top layers. This observed symmetry in axial resolution indicates that there is not much degradation of the axial resolution over the B-scan plane. The spatial distribution of axial and lateral resolutions shows that the axial resolution is uniform across the B-scan, where as the lateral resolution varies significantly with axial position.

Although the size of the structures in the phantom are larger than the resolution of the system, the results show that it is possible to measure the spatially varying PSF of the system using these phantoms. The obtained values of the resolutions may not reflect the true resolution of the system, however they

demonstrate the feasibility of the quantitative characterisation method.

The fabricated structures need to be smaller than $3 \mu m$ to experimentally measure the PSF of an OCT system with a resolution of $8 \mu m$ (FWHM). The parameter space studies indicate that it is possible to create structures of less than $2 \mu m$ ¹. The structures in the PSF pattern were inscribed at relatively higher energy, thus producing large structures that can be easily identified during our studies. The structures inscribed in the PSF pattern have a lateral dimension, measured using a microscope, of around $4.1 \mu m$, thus the effect of PSF broadening is noticeable in the observed data. The structures inscribed at relatively high energy exhibit some asymmetry. For an OCT with a resolution of $8 \mu m$, the expected measured size of the structures would be around $14.19 \mu m$.

In this study, data from the PSF pattern was used for PSF and MTF calculations. The distortion pattern can also be used for resolution measurements (PSF and MTF) and sensitivity drop-off. The volumetric data from the PSF pattern obtained using the MD-OCT system is used to demonstrate the quantitative characterisation methods that can be used to estimate the spatially variant PSF. Although, due to the size of the structure, some PSF broadening effects will be observable, these will not influence the quantitative method described in this section.

7.1.3 Summary

This section described the quantitative methods used to measure the two dimensional resolution of an OCT system using an OCT-phantom. The technique demonstrated here can be extended to three dimensional with ease. The structures in the single layer phantoms are much smaller than the OCT resolution hence PSF can be directly measured. Whilst the single layer artefact is limited to characterisation around the focal point of the system, tilting the sample perpen-

¹New OCT-phantoms were fabricated with lower energy (i.e. 30% and 20%) of the total energy. Latest PSF samples (inscribed with energy of 30%(55 mW) and translating the sample by 10 mm/s produced structures with a lateral width $< 1.8 \mu m$.

dicular to the scan plane allows the system resolution away from the focal point to be characterised.

Ultimately, this proposed solution has the potential to be extended to deconvolve the object function from its PSF and render the data in real time.

7.2 Lateral resolution

The lateral resolution pattern can be used to qualitatively identify the resolution limit of the system. It is also possible to quantitatively estimate the lateral resolution of the OCT. To make a qualitative comparison, data were obtained from two OCT systems: MD-OCT and Santec-OCT.

Using the MD-OCT, an area of $4 \text{ mm} \times 0.8 \text{ mm}$ was imaged, which covers the set 1-5 (i.e. line spacing of $1 - 50 \mu\text{m}$). Volumetric data consist of 200 B-scans each separated by $4 \mu\text{m}$, whilst each B-scan is composed of 962 A-scans. In the MD-OCT system, the sample was tilted by 1.22° , in relation to the beam axis, in order to minimise the strong back reflection by its surface. In the Santec-OCT system, an area of $2 \text{ mm} \times 2.5 \text{ mm}$ was also imaged, and 500 B-scans were obtained with a spacing of $4 \mu\text{m}$.

Figure 7.4 shows the B-scans of the lateral resolution pattern using both OCT systems. MD-OCT system has an axial and lateral resolution of $11.4 \mu\text{m}$ and $8.4 \mu\text{m}$ respectively. Santec-OCT system has an axial and lateral resolution of $8 \mu\text{m}$ and $23.1 \mu\text{m}$ respectively.

From B-scans we can clearly see that the MD-OCT, exhibits better resolution than the Santec-OCT. Also, the images clearly show the degradation of lateral resolution with depth. The drop in lateral resolution with depth is clearly visible from these data, where the spacing between the points diminish and start to disappear with depth as well as the points start to blur and spread out more with increasing depth. Because the imaging depth of the MD-OCT system (of each channel) is about $250 \mu\text{m}$ and the depth of the structure is about $600 \mu\text{m}$ which is

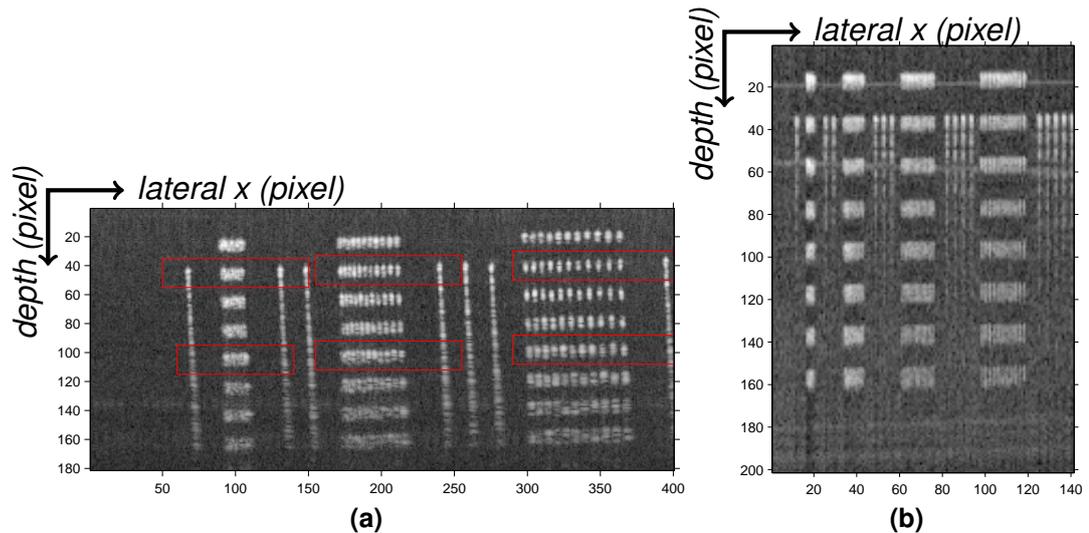


Figure 7.4: B-scan data of the lateral resolution pattern obtained by a) MD-OCT system and b) Santec-OCT system. Both images are presented in log scale and scale are in pixels. Santec-OCT system (b) has lower resolution, hence less number of pixels.

well outside the focal distance of a channel².

Resolution limit of the system can be assumed to be equal to the line spacing that is least resolvable by the system. Approach here uses data from the MD-OCT system. A small region around each row of points were selected to form an *en-face* image of that layer. *En-face* data is formed by summing the intensities selected in the sub-region along the z-axis (beam axis). Figure 7.5 shows the *en-face* images (top) and oversampled line profile (bottom) extracted from set 1 (left) and 2 (right) of the lateral resolution pattern. The *en-face* images correspond to the area marked by red rectangles in Figure 7.4a (i.e set 1 and 2 from second layer).

²Data from only one channel was used.

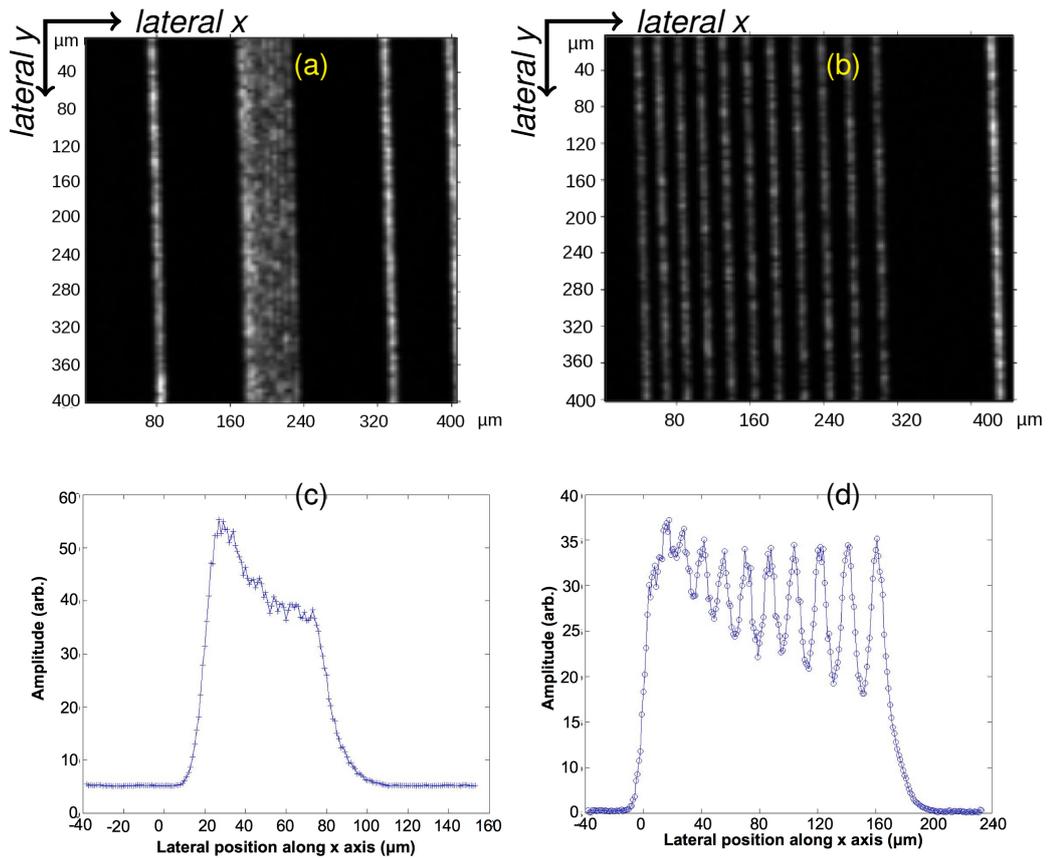


Figure 7.5: *En-face* images (top) and oversampled line profile (bottom) extracted from set 1 (left) and 2 (right) of the lateral resolution pattern.

Each raw data set in the *en-face* image is essentially a lateral line profile. The peak position of the index lines, in the *en-face* images, were estimated by fitting a one-dimensional Gaussian function. Each line profile shifted laterally, so the peaks of the index (right hand) are overlaid on top of each other.

Line profiles of set 1 and 2 (spacing of 1–10 μm and 11–20 μm) were extracted from the *en-face* image. Lines from each set were oversampled by registering all corresponding lines to a common coordinate and then oversampled four times the pixel width as shown in Figure 7.5c-d. Results indicate that the separation between peaks increases with increasing spacing, which could be used to qualitatively measure the resolution limit.

7.3 Sensitivity

Sensitivity of MD-OCT system had previously been determined using a conventional method of measuring the reflectance of a known reflector. The reflector for the measurement of sensitivity used was a combination of a tilted reflective neutral density filter in combination with a mirror³. The mirror reflectance and filter transmittance have been accurately calibrated against national standards. The mirror was aligned normal to the OCT measurement beam and carefully positioned at the focus of the instrument by maximizing the measured signal intensity of the reflector. Several measurements were made to determine repeatability and to assess the uncertainty. The limiting uncertainties were found to be the noise floor and the spectral flatness of the transmission filters.

The faintest structure detected by the OCT system was the first pair. The previously measured signal to noise ratio of the OCT system is 90.1 ± 0.9 dB (95% confidence).

This work demonstrates the potential of using a fs inscribed phantom to evaluate the sensitivity limit of an OCT system. By inscribing varying reflectivity structures inside a transparent material (i.e. Fused silica) using fs pulses, a three dimensional sensitivity phantom can be fabricated. The results indicate that the structures with varying reflectivity have the potential of measuring the sensitivity of an OCT system as a function of lateral and axial distance. Most importantly, these artefacts can be used to monitor the system performance over time.

³This work was done by P.D.Woolliams

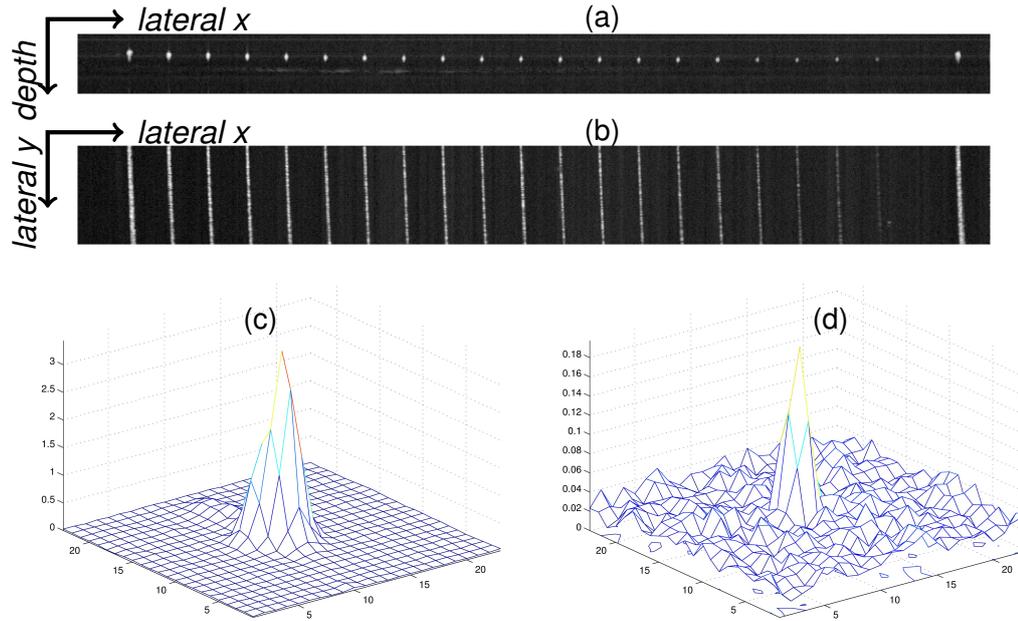


Figure 7.6: OCT images of structures inscribed with varying energy at $100 \mu\text{m}$ below the surface of fused silica used as sensitivity phantom. a) B-scan and b) en-face image of the sample that was used for qualitative sensitivity test. Mesh plots c-d shows the profile of the first and the one before last point (faintest visible) (i.e. last point is marker inscribed at higher energy).

Lines and cross sectional profiles of structures, in *en-face* and B-scan data, can be approximated using one- and two-dimensional Gaussian functions as follows:

$$G_{1d}(x; \lambda_{1d}) = Ae^{\left(-\frac{(x-x_0)^2}{\omega_x^2}\right)} \quad (7.3.1)$$

$$G_{2d}(x, y; \lambda_{2d}) = Ae^{\left(-\frac{(x-x_0)^2}{\omega_x^2} - \frac{(y-y_0)^2}{\omega_y^2}\right)} + g_0 \quad (7.3.2)$$

where λ refers to the set of Gaussian fit parameters $\{A, x_0, y_0, \omega_x, \omega_y, k\}$, with A the amplitude, x_0 and y_0 the coordinates of the centre of the peak, ω_x and ω_y the $1/e$ axial and lateral width of the peak and k the peak offset. Integrated intensity of the structures can be calculated by computing the integration of the estimated Gaussian peak as follows:

$$\int \int G_{1d}(x; \lambda_{1d}) dx = -\frac{1}{2} A \sqrt{\pi} \omega_x \operatorname{erf} \left(\frac{x - x_0}{\omega_x} \right) \quad (7.3.3)$$

$$\int \int G_{2d}(x, y; \lambda_{2d}) = \frac{1}{4} A \pi \omega_x \omega_y \operatorname{erf} \left(\frac{x - x_0}{\omega_x} \right) \operatorname{erf} \left(\frac{y - y_0}{\omega_y} \right) \quad (7.3.4)$$

Figure 7.7 shows the integrated intensity of the lines and cross section, inscribed with decreasing energy, approximated using the equation (7.3.4). Results were averaged over 100 frames. A third order polynomial was fit and 95% confidence bond to these estimates are also shown in the Figure.

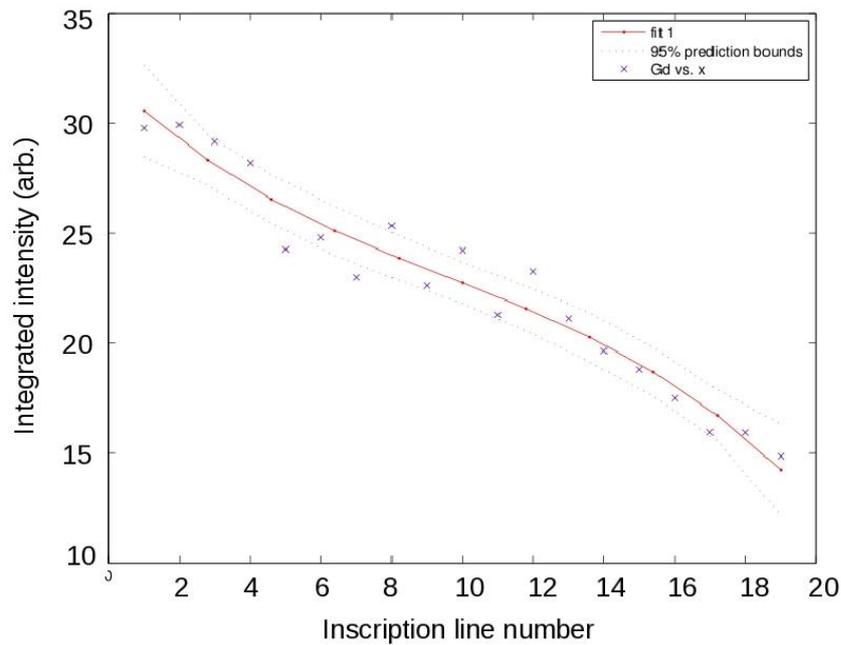


Figure 7.7: *Integrated intensity from sensitivity sample.*

In conclusion, structures with varying dimension and reflectivity were fabricated using fs laser inscription, for characterizing the sensitivity of an OCT system. The smallest reflective point the OCT system can resolve determines the sensitivity limit of the system, giving an easy method of monitoring the system performance.

7.4 Sensitivity drop

The PSF pattern can be used to measure the relative sensitivity drops of the OCT system. Figure 7.8 shows the B-scan of the PSF pattern and sensitivity drops estimated from the volumetric data of the PSF pattern. Integrated intensity of the structures were estimated as described previously for all the detected points in each B-scan. Values were then averaged over 25 B-scans and interpolated across the B-scan image plane. Results show the drop in sensitivity away from the focal point.

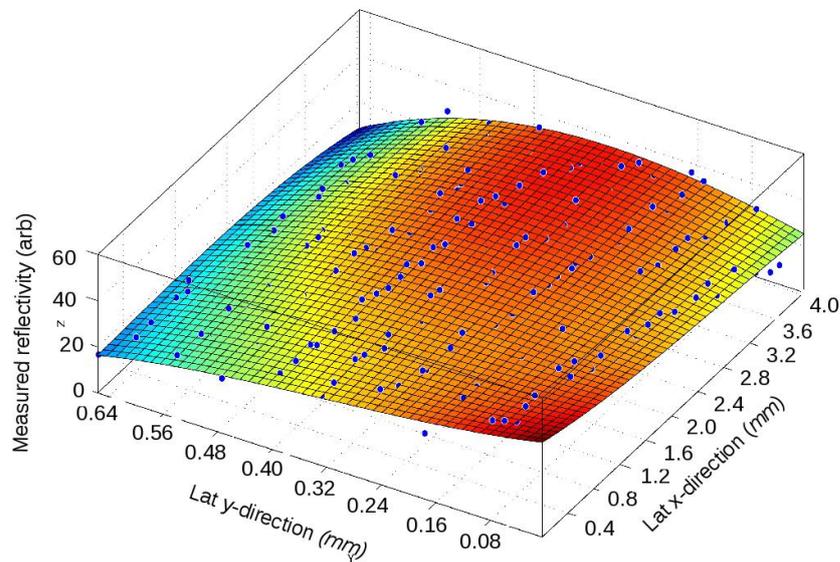


Figure 7.8: a) The B-scan of the PSF pattern and b) the measured intensity of the points across the B-scan plane.

We are currently refining the inscription parameters and phantom design to improve usability. A comprehensive analytical method is also under development, which will allow the operator to quantify the sensitivity performance of the OCT system.

7.5 Distortion

Figure 7.9 shows an identified intersection and measured distortion tensor (a) and a magnified 2 dimensional subsection (b). The data was obtained from the distortion phantom using a Michelson diagnostic OCT system. The intersections were identified manually using an iterative procedure and the distortion tensor was calculated by subtracting the expected reference point from the measured intersection. This simple procedure illustrates the technique of measuring the three dimensional distortion of an OCT system using the grid patterns. Methods that automatically and more accurately locate the intersections and carry out the distortion measurement and correction procedure are in development.

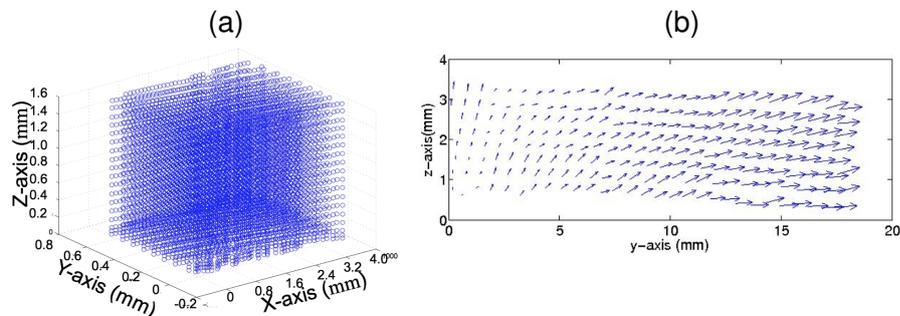


Figure 7.9: Identified intersections (a) and measured distortion tensor of part of the select points (b) from the sample data set.

The offset observed in Figure 7.9a arise due to the tilt in the sample.

7.6 Conclusion

This chapter briefly described the characterisation of an OCT system using our OCT-phantom. We used singled layered PMMA phantom and a multilayred phantom, that contains four patterns for evaluating the OCT system. Spatially varying PSF, sensitivity and distortion of the OCT can be accurately measured using the OCT-phantoms. Future work will involves developing better refined quantification methods/algorithms.

Chapter 8

Deconvolution

Following section describes the image enhancement by deconvolution using the spatially varying PSF measured using OCT-phantoms. Spatially varying PSF presented in section §7.1.2 were used. In this study, overlap-add method is used for deconvolution. Segmented data is zero-padded and deconvolved with the corresponding PSF. Overlapping regions were added together to form the complete deconvolved B-scan image.

8.1 Method

Data of the grid pattern obtained using the MD-OCT system were used as the test data for deconvolution. This makes it easy to quantitatively analyse the resolution improvement by estimating the FWHM axial and lateral width of the points by fitting a two dimensional Gaussian function before and after deconvolution.

A B-scan was segmented into 50×50 pixel grid, and the average of the PSF corresponding to each tile was estimated from the PSF-map as follows:

$$\{G(x, y)\}_{(m,n)} = G_{2D}(x, y; \langle B_{m,n} \rangle) \quad (8.1.1)$$

$$\langle B_{m,n} \rangle = \sum \frac{1}{N} B_{m,n}^i \quad (8.1.2)$$

CHAPTER 8. DECONVOLUTION

where the indices of the sub-regions are denoted by $m = [0, M)$ and $n = [0, N)$ with $M = \lceil \frac{X}{\Delta X} \rceil$ and $N = \lceil \frac{Y}{\Delta Y} \rceil$ where $X, Y, \Delta X$ and ΔY refer to the sizes of the image and of the sub region. The fit parameters estimated for the regions (m, n) are denoted by $\langle B_{m,n} \rangle$.

All PSF values lying within each sub-section were averaged to represent a single PSF of that region, where it was assumed that the PSF of the sub-region is spatially invariant. In the illustration in Figure 8.1, four measured PSF points ($P_1 - P_4$) lie in the region $R(n, m)$. The average PSF of this region is given by $P_{av} = (1/4) \sum_{i=0}^4 P_i$. Average of these PSF in each tiles were used to represent the PSF of that region. For those regions with no measured PSF (i.e. top corners) average of the neighbours was used. Here we are only interested in the 2D deconvolution of the B-scans, hence the z-axis is not considered.

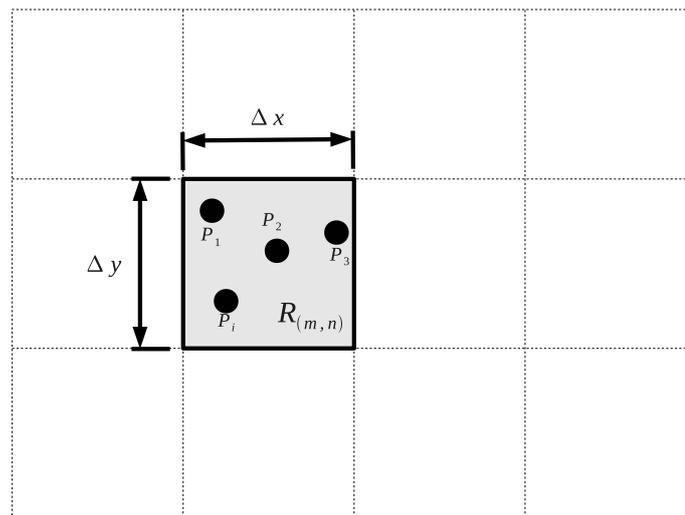


Figure 8.1: Example of calculating average PSF of a tile.

The axial and lateral widths of the structures after deconvolution were also estimated by fitting a Gaussian function and the statistical distribution is presented as histogram. Improvements of the deconvolution were quantitatively analysed by estimating the FWHM axial and lateral width of the points in the B-scan data used for deconvolution. Observed resolution improvement after deconvolution were calculated by computing the difference between the mean distribution before and after deconvolution.

8.2 Results

8.2.1 Blurred image

Figure 8.2 shows the B-scan of the distortion phantom. The blue rectangle indicates the region where from spatially varying PSF were obtained using the PSF phantom in section §6.2.2. The tiles are indicated by the red lines and the green rectangle indicates small region of the B-scan that will be used for qualitative comparison.

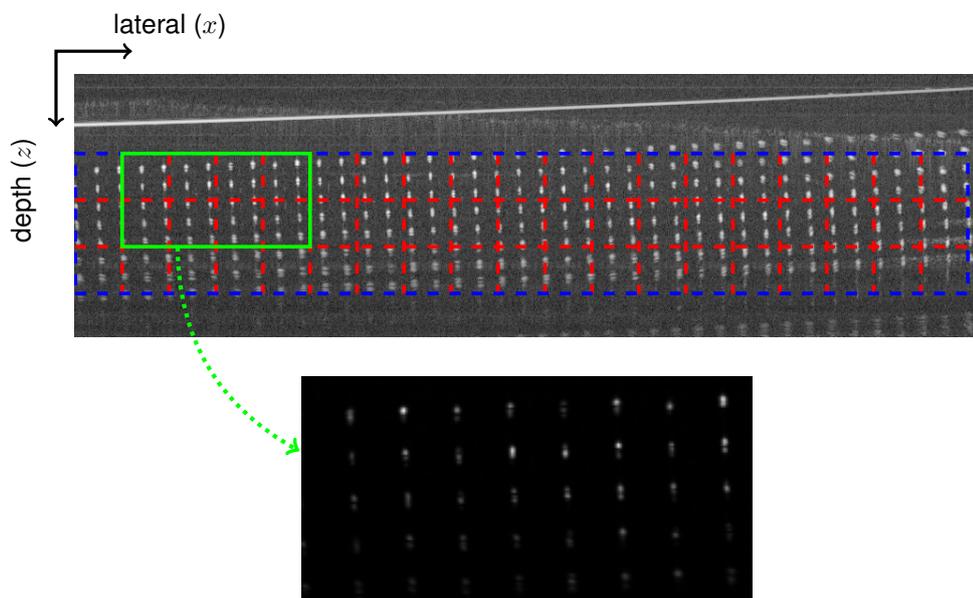


Figure 8.2: *B-scan of blurred image (top) and region used for qualitative comparison. Blue rectangle represent the region that can be deconvolved using the spatially PSF data obtained using PSF pattern, which is split into tiles indicated by red rectangles. The green rectangle represent the region (show in bottom) that used for qualitative comparison.*

ROI, indicated by the blue rectangle in Figure 8.2, was split into 3×19 tiles, each tile composed of 50×50 pixels, marked by red lines in B-scan. Each tile was deconvolved using their corresponding PSF using WF and LR methods. Results from blurred (original) and improvements after applying LR and WF deconvolution methods are presented in this section.

Figure 8.3 shows the statistical distribution (histogram) of the axial and lateral widths of the points prior to deconvolution (i.e. in the blurred image). The original

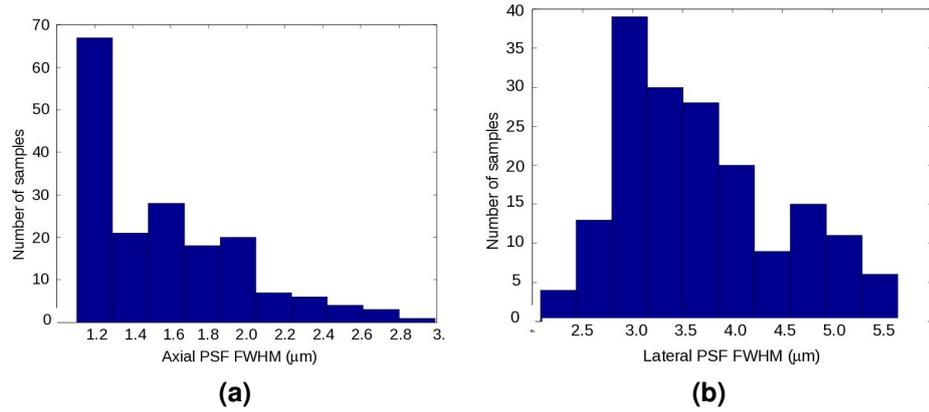


Figure 8.3: Histograms of (a) axial and (b) lateral resolution prior to deconvolution.

(blurred) data has a mean (model) axial and lateral resolution of $14.74 \pm 0.567 \mu m$ and $34.94 \pm 1.150 \mu m$ respectively.

Table 8.1: The summary of quantitative measure of the deconvolved tiled mean PSF

LC	15.73 ± 0.50
CNR	24.75 ± 0.14
average SNR	21.29 ± 0.23
sharpness (K)	1.51 ± 0.12

8.2.2 Mean PSF

Each tile was deconvolved using average PSF calculated over the entire B-Scan. Figure 8.4 shows PSF corresponding to each tile. In Figure 8.5, first and second rows shows results corresponds to LR and WF deconvolution. B-scans (first column) and mesh plots (second column) qualitatively indicate the improvements after deconvolution. Whilst last two column show axial and lateral FWHM. Histograms qualitatively indicate the improvements after deconvolution. Results show axial and lateral resolution having a model resolution of $9.0738 \pm 3.5259 \mu m$, $22.6640 \pm 8.7257 \mu m$ for LR routine and $11.3937 \pm 3.5742 \mu m$, $26.8024 \pm 7.9057 \mu m$ for WF routine respectively. This is equivalent to an improvement of $(5.6655 \mu m, 12.2778 \mu m)$ for axial resolution and $(3.3456 \mu m, 8.1394 \mu m)$ for lateral resolution.

Table 8.2 summaries the resolution improvement performance matrices LC ,

CNR , SNR_{av} and sharpness(K). These were obtained from the histograms shows in Figure 8.6.

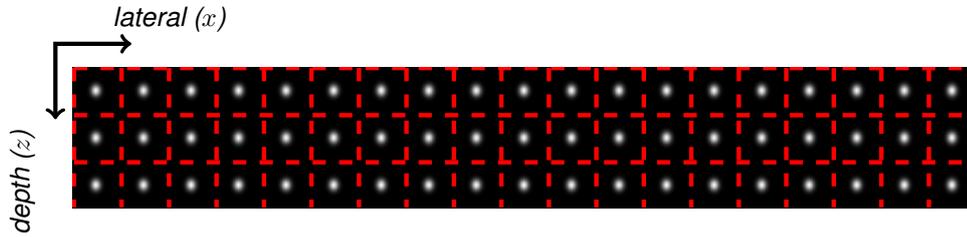


Figure 8.4: Mean PSF used for deconvolution of each sub section.

Table 8.2: The summary of quantitative measure of the deconvolved tiled mean PSF

	LR	WF
LC	123.07 ± 6.39	29.83 ± 1.03
CNR	31.54 ± 0.30	27.64 ± 0.27
average SNR	31.0 ± 0.32	25.90 ± 0.31
sharpness (K)	9.50 ± 0.65	2.90 ± 0.054

8.2.3 Varying (Local) PSF

The two dimensional Gaussian function of the corresponding section $G_{i,j}$ were generated using the corresponding axial w_x^{ij} and lateral w_y^{ij} FWHM resolution. Each sub-region was then deconvolved using the corresponding PSF.

Figure 8.7 shows the PSF corresponding to each tile. In Figure 8.5, first and second rows shows results corresponding to LR and WF deconvolution. B-scans (first column) and mesh plots (second column) qualitatively indicate the improvements after deconvolution. Whilst last two columns shows axial and lateral FWHM. Histograms qualitatively indicates the improvements after deconvolution. Results show axial and lateral resolution having having a model resolution of $9.4720 \pm 3.7396 \mu m$, $23.0140 \pm 8.1706 \mu m$ for LR routine and $11.8459 \pm 4.9130 \mu m$, $28.0137 \pm 7.1809 \mu m$ for WF routine respectively. This is equivalent to an improvement of $(5.2672 \mu m, 11.9277 \mu m)$ for axial resolution and $(2.8934 \mu m, 6.9280 \mu m)$ for lateral resolution.

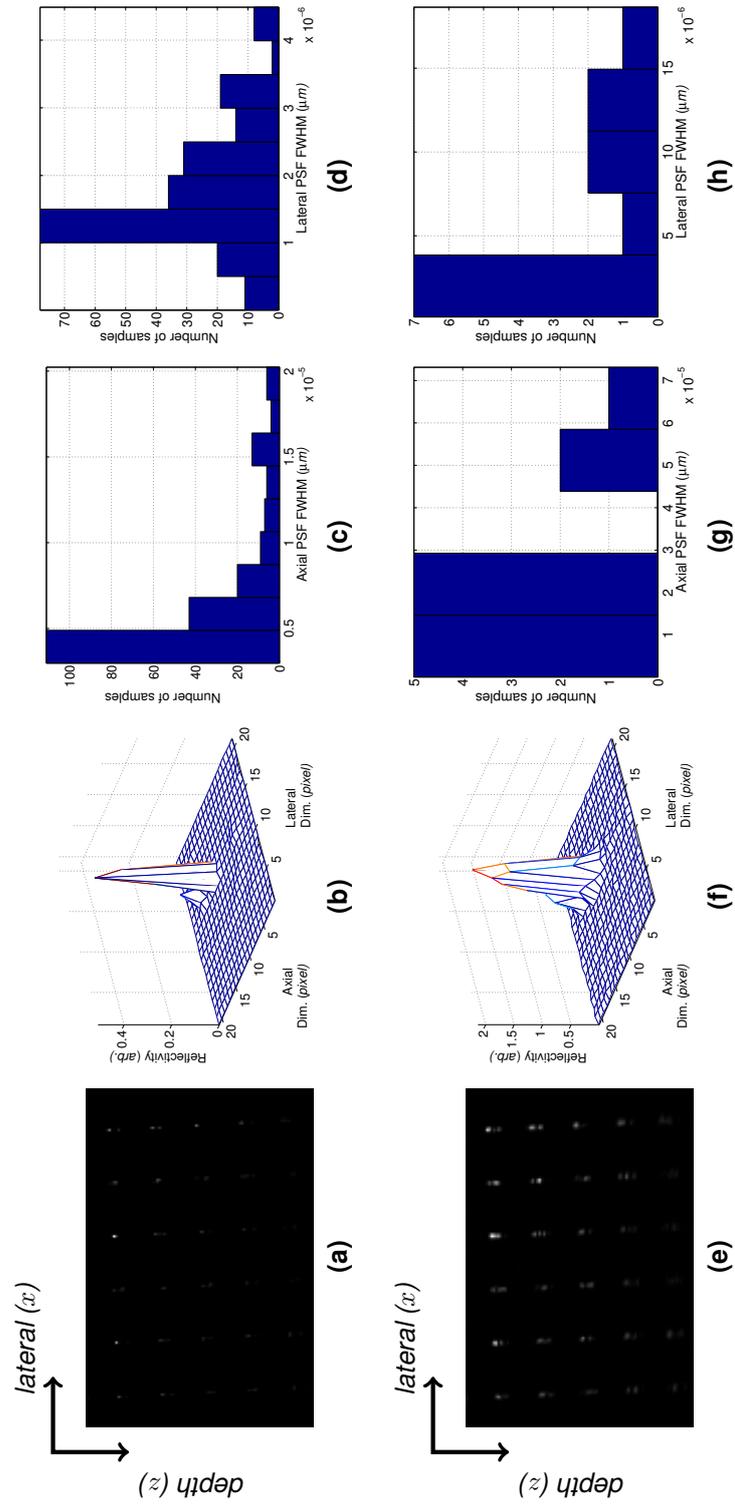


Figure 8.5: The results after applying deconvolution using LR (top row) and WF (bottom row) routines using mean PSF. B-scan after deconvolution (a,e); mesh plots (b,f); axial FWHM (c,g); lateral FWHM (d,h).

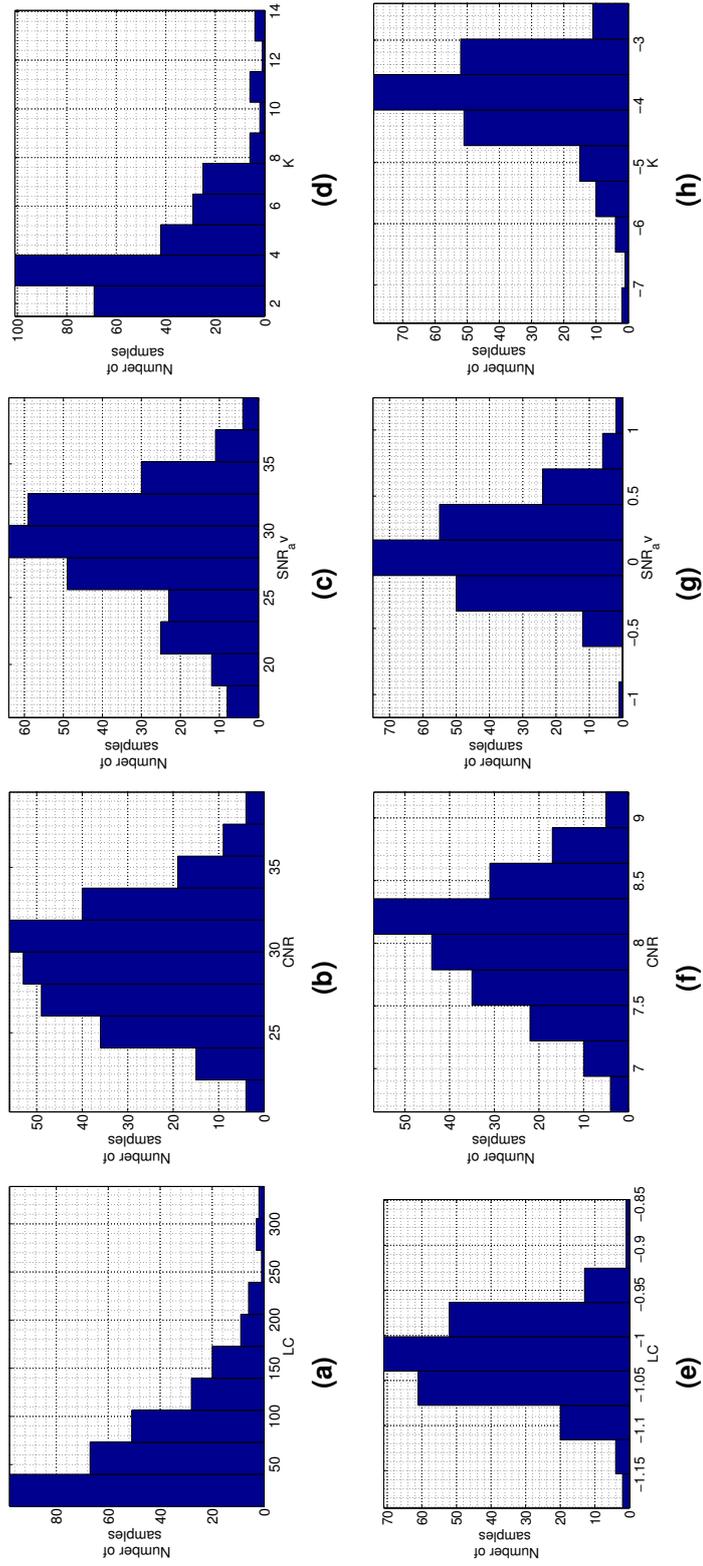


Figure 8.6: Statistical distribution of $LC(a,e)$, $CNR(b,f)$, SNR (c,g) and sharpness $K(d,h)$ for LR(top row) and WF (bottom) routines using mean(global) PSF.

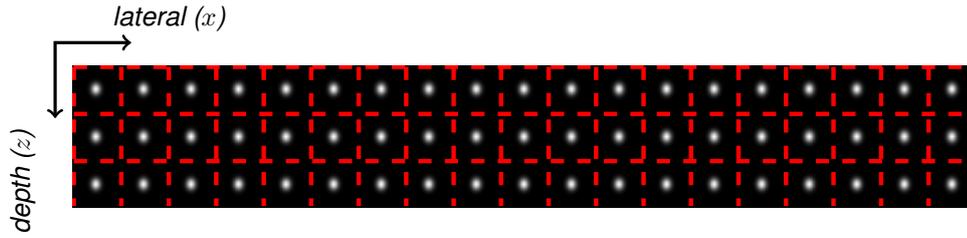


Figure 8.7: Local (mean) PSF corresponding to each tile, that was used for deconvolving that tile. PSF within each tile is assumed to be invariant.

Table 8.3 summarizes the resolution improvement performance matrices LC , CNR , SNR_{av} and sharpness(K). These were obtained from the histograms shows in Figure 8.9.

Table 8.3: The summary of quantitative measure of the deconvolution Tiled Local PSF.

	LR	WF
LC	106.51 ± 4.80	29.32 ± 1.04
CNR	30.65 ± 0.29	27.25 ± 0.26
SNR_{av}	30.04 ± 0.32	25.54 ± 0.32
K	14.52 ± 5.21	2.84 ± 0.05

8.3 Discussion

In both LR and WF cases, the points appear to be sharper and the FWHM axial and lateral width of the point decreases (i.e. resolution improves) after applying deconvolution, that is apparent from the mesh plots shown in the Figures 8.5(b,f)-8.8(b,f). These results leads to qualitatively confirming the increment in the resolution.

Statistical distribution of the quantitative measurements of the image enhancements are shown in the histogram plots in Figure 8.6 and 8.9.

Quantitative analysis also confirms the improvements in the quality of the image after deconvolution and the results indicate that overall LR achieves better results than the WF routines. The WF algorithm heavily depends on the signal to noise ratio of the system, and hence a wrong estimate of SNR can lead to inaccurate results. In this study, the WF was used for demonstration purposes.

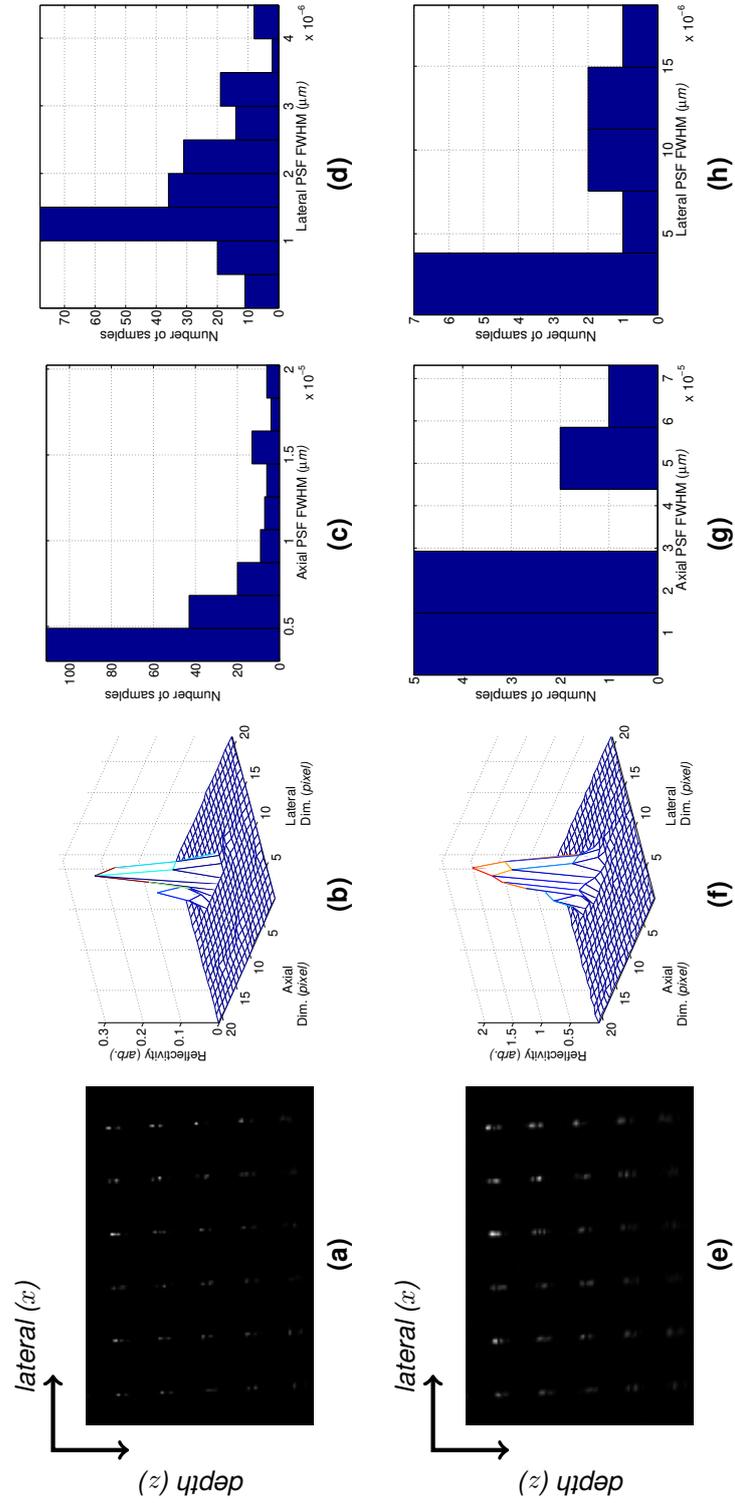


Figure 8.8: The results after applying deconvolution using LR (top row) and WF (bottom row) routines using tiled PSF (i.e. PSF is locally invariant within each tile). B-scan after deconvolution (a,e); mesh plots (b,f); axial FWHM (c,g); lateral FWHM (d,h).

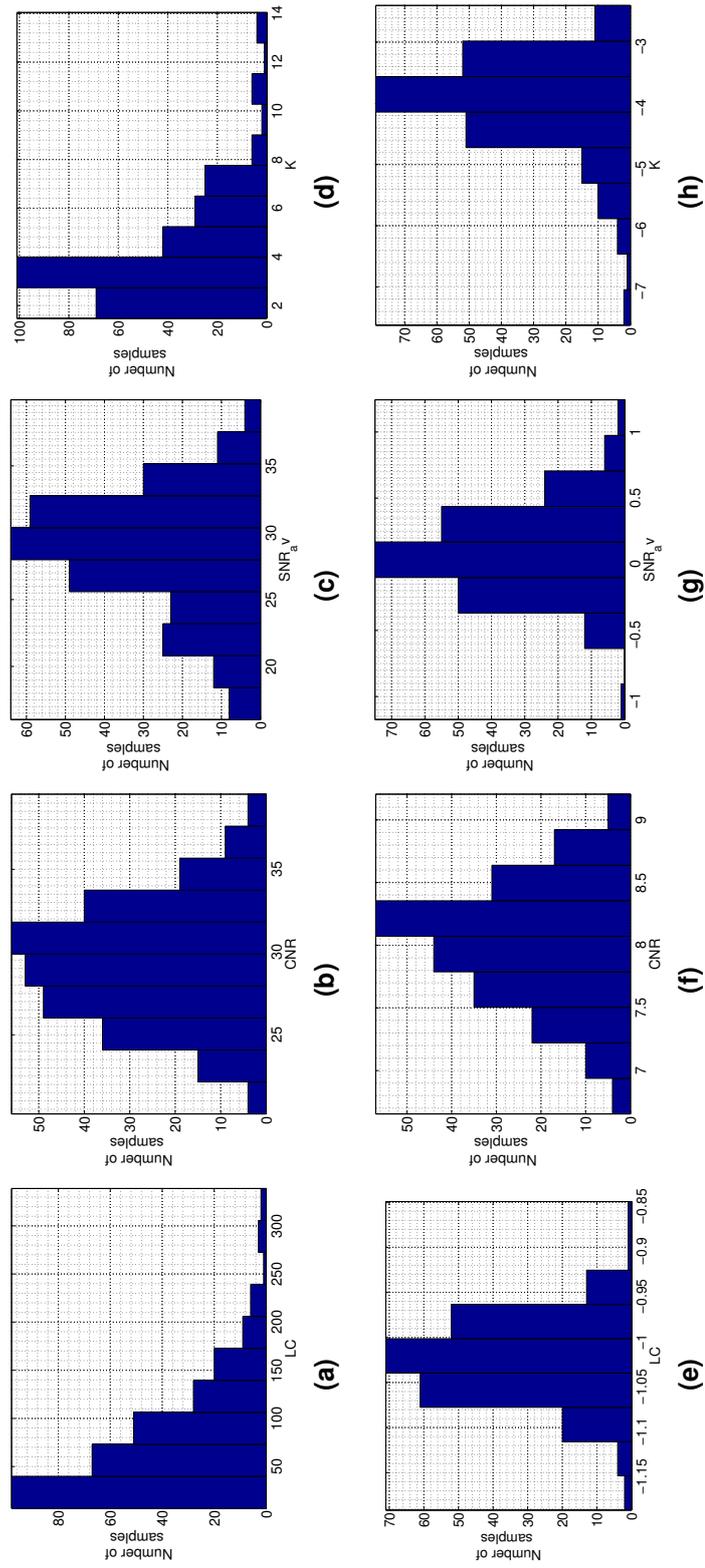


Figure 8.9: Statistical distribution of LC(a,e), CNR(b,f), SNR(c,g) and sharpness K (d,h) for LR(top row) and WF (bottom) routines using local PSF

An accurate measurement of noise of the system is required in order to realize the full potential of the deconvolution method. This is left as a future study.

Other factors such as dispersion mismatch between the sample and reference arms, speckle noise, and multiple scattering, can also contribute to the blurring effect observed in the OCT. Removal of the blur by deconvolution only accounts for the effects caused by PSF. This limits the sharpness and the contrast of improvement that can be realised by applying deconvolution. Deconvolving the images could reveal blurring caused by factors other than PSF.

8.4 Conclusion

Improvement in the resolution of OCT data by deconvolution using experimentally measured spatially varying PSF is demonstrated. Both axial and lateral resolution of the OCT can be deconvolved using a 2D PSF with the assumption that the PSF in each sub-region is spatially invariant.

Chapter 9

Data Processing

This Chapter will discuss the optimisation of the GPU based processing method, exploiting several of the available resources and GPU based rendering that avoids copying the data back to CPU, to achieve nearly $1.8 \text{ millionlines/s}$. This was first published in 2011 and has been further developed since then [101]. Progress to date (at the time of writing) has not explicitly exploited the different GPU memory types available or combined both GPU processing and rendering capabilities to increase efficiency and speed. Here we employ the use of paged memory to increase the data transfer rate from the CPU (host) to the GPU (device). Furthermore, OCT image data is rendered on the GPU to increase overall speed, rather than copying the data back to the CPU. All algorithms are optimized for the specific GPU used.

9.1 Interpolations

In SD-OCT, the depth profile of the sample is obtained by applying a discrete FT to the detected intensities. Generally in SB-OCT, a grating is used to disperse the wavelength components almost linearly in wavelength and detected by an array detector. In SS-OCT the laser is swept linearly in wavelength, therefore the detected signal is evenly sampled in wavelength [147].

An A-scan can be reconstructed by taking the fast Fourier transform (FFT) of the detected spectrum. The application of discrete Fourier transform (DFT) by FFT requires the data to be uniformly sampled in wave-number space. However, the detected signal is uniformly sampled in wavelength, hence the sample data is non-uniformly distributed in wave-number space.

Taking an FFT of unevenly distributed data leads to the broadening of the coherence envelope and the sensitivity falls off with depth. Therefore, the data has to be resampled to achieve a uniform distribution in the k - space in order to apply an FFT [9]. Resampling from λ to k is commonly calculated by interpolating the values measured that are sampled uniformly in λ space to predefined calibration values uniformly sampled in k space [100]. The interpolation used for resampling is a computationally intensive process in OCT data processing.

The accuracy of the resampling technique used is important for the reconstruction, since the efficiency of the interpolating algorithm tends to degrade as the sampling rate approaches the *Nyquist limit*. Errors in the resampling cause a drop in the sensitivity with depth. Linear and cubic-spline interpolations are conventionally used for resampling the data. Linear interpolation is the simplest method, in which interpolations are calculated from the two nearest points using a first order linear equation. This simple method suffers from large interpolation errors as the data reaches Nyquist limit. Cubic spline interpolation uses a piecewise polynomial function to interpolate data between two known points. The accuracy of the cubic interpolation is better than linear interpolation but requires longer processing times.

Several methods have been developed to reduce the interpolation errors and improve the sensitivity fall-off by using additional hardware or processing algorithms. Hu and Rollins [98] proposed a linear-in-wavenumber spectrometer, that used a custom made prism to disperse the spectral components linearly in wavenumber, which eliminated the need for resampling. Zhang et. al. [91] proposed a time-domain interpolation method. This method is relatively slow, but improves the sensitivity fall-off by 2 dB. Wang et. al. [148] showed improvement in the sensitivity fall-off, by adopting non-uniform discrete Fourier transform to compute the DFT by direct matrix multiplication.

In this works, a linear interpolation and a third order Lagrangian interpolation are implemented on the CPU and the GPU. The following sections describe the implementation of the algorithms and give the benchmark results. Here and throughout this thesis, CPU results are taken as the gold standard, against which the GPU results are evaluated.

9.1.1 Linear interpolation

Linear interpolation is the simplest re-sampling technique used to determine the value of y (i.e. intensity) at a given unknown coordinate x (i.e. pixel position) between the interval $[x_1, x_2]$, given the values at two points, (x_1, y_1) and (x_2, y_2) , using the equation:

$$y = y_1 + (x - x_1) \frac{y_2 - y_1}{x_2 - x_1}, \quad (9.1.1)$$

9.1.2 Lagrangian Interpolation

The polynomial of the Lagrangian interpolation $P(x)$ of degree n that passes through $n + 1$ points $(x_i, y_i)_{i=1}^n$ where $x_i \neq x_j$ is given by:

$$P(x) = y_i \sum_{j=1}^n P_j(x), \quad (9.1.2)$$

CHAPTER 9. DATA PROCESSING

where y_i is the intensity at the point x_i and x is the resample value and $P_j(x)$ is the Lagrangian basis polynomial given as:

$$P_j(x) = \prod_{k=1, k \neq j}^n \frac{x - x_k}{x_j - x_k} \quad (9.1.3)$$

In this work, a third order Lagrangian interpolation polynomial is used for resampling the data. A third order polynomial requires four points around the interpolating point r_i . Interference data are recorded by a linear detector (i.e. pixels are linear in space), so the indices of the four nearest pixels (two on either side of the interpolating point) were used. Figure 9.1 illustrates the resampling point and the four points (pixel indices) used. For each resampling point r_i four points in the range $[\lfloor r_i \rfloor - 1, \lceil r_i \rceil + 1]$ were chosen, where $\lfloor r_i \rfloor$ is the greatest integer smaller than r_i and $\lceil r_i \rceil$ is the smallest integer larger than r_i , which can be expressed mathematically as:

$$\lfloor r_i \rfloor = \min\{m \in \mathbb{Z} | m \leq r_i\} \quad (9.1.4)$$

$$\lceil r_i \rceil = \max\{n \in \mathbb{Z} | n \geq r_i\} \quad (9.1.5)$$

Therefore x_j where $i = \{m \in \mathbb{Z} | m \leq 4\}$ can be written as:

$$\begin{aligned} x_1 &= \lfloor r_i \rfloor - 1 \\ x_2 &= \lfloor r_i \rfloor \\ x_3 &= \lceil r_i \rceil \\ x_4 &= \lceil r_i \rceil + 1 \end{aligned} \quad (9.1.6)$$

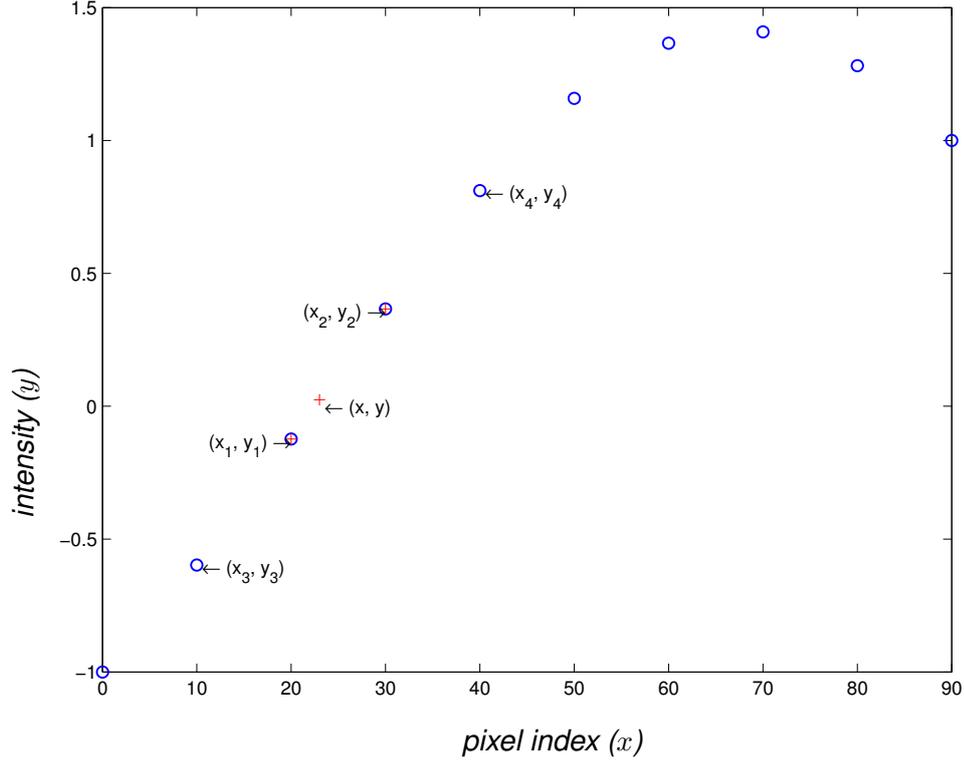


Figure 9.1: Illustrating the Lagrangian interpolation method. Point (x, y) represents the resampling location whilst points (x_i, y_i) corresponds to where $i = 1 - 4$ points either side of the interpolating points used for computations.

By substituting equation (9.1.6) into equation (9.1.3) the interpolation coefficients can be explicitly written as follows:

$$P_{1,i} = \frac{(r_i - \lceil r_i \rceil)(r_i - \lfloor r_i \rfloor)(\lceil r_i \rceil - r_i + 1)}{(\lceil r_i \rceil - \lfloor r_i \rfloor + 1)(\lceil r_i \rceil - \lfloor r_i \rfloor + 2)}, \quad (9.1.7)$$

$$P_{2,i} = \frac{(\lceil r_i \rceil - r_i)(\lceil r_i \rceil - r_i + 1)(r_i - \lfloor r_i \rfloor + 1)}{(\lceil r_i \rceil - \lfloor r_i \rfloor)(\lceil r_i \rceil - \lfloor r_i \rfloor + 1)}, \quad (9.1.8)$$

$$P_{3,i} = \frac{(r_i - \lfloor r_i \rfloor)(\lceil r_i \rceil - r_i + 1)(r_i - \lfloor r_i \rfloor + 1)}{(\lceil r_i \rceil - \lfloor r_i \rfloor)(\lceil r_i \rceil - \lfloor r_i \rfloor + 1)}, \quad (9.1.9)$$

$$P_{4,i} = \frac{(r_i - \lceil r_i \rceil)(r_i - \lfloor r_i \rfloor)(r_i - \lfloor r_i \rfloor + 1)}{(\lceil r_i \rceil - \lfloor r_i \rfloor + 1)(\lceil r_i \rceil - \lfloor r_i \rfloor + 2)}, \quad (9.1.10)$$

where the subscript i represents the index of the resampling table.

9.2 Method

Figure 9.2 shows the simplified standard processing flow chart for the CPU-GPU hybrid model. When the program starts, it reads the header information such as the resample table and the background light from a file.¹ At the initialisation stage, the program searches for a compatible GPU. If compatible GPUs are present on the PC then by default the highest performance GPU is initialised for data processing. It is possible to explicitly select a GPU from the list and also possible to run the processing on multiple GPUs. During the GPU initialisation stage, memory required for data processing is allocated on the GPU, header data are transferred to the GPU memories (i.e different GPU memories are used for different variables). Data is transferred to the GPU and the standard processing is computed on the GPU. In the absence of a compatible GPU, the processing takes the alternative paths indicated by the dot-dot lines and the data is processed by the CPU.

Figure 9.3 shows the processing flow chart for validating the GPU results. The standard OCT data processing is split into three parts. First the pre-processing stage, which involves the background subtraction and interpolations of the data from linear in wavelength to linear in wavenumber using either linear or Lagrangian interpolation. Then the Fourier transform of the pre-processed data is computed using either Fastest Fourier Transform in the West (an FFT library) (FFTW) or CUDA fast Fourier transform (CUFFT) on the CPU or GPU respectively. Finally, the modules are computed to transfer the data from the Fourier space to real space and converted to dB . In addition to these processes, data need to be transferred to the GPU. In the case of real time rendering, copying back to CPU can be avoided by directly rendering data from the GPU.

¹The data were acquired on by EX1301 Michelson SS-OCT and stored as binary files. Data such as background were extracted from data acquisition (DAQ) in real time. The header file contains the resampling table, background, stray light and other information about the system and the data

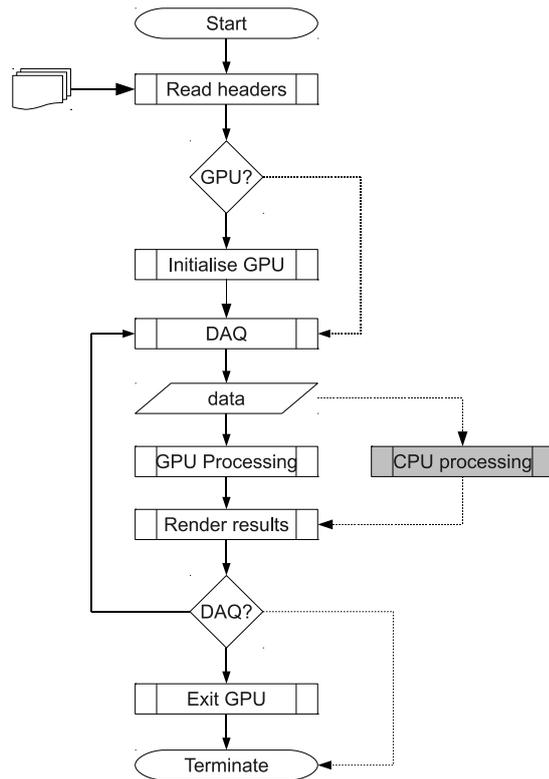


Figure 9.2: Schematics of the GPU based data processing flowchart. The default paths are marked by bold lines. When the default path fails, the processing takes via the alternative path indicated by the dot dot lines.

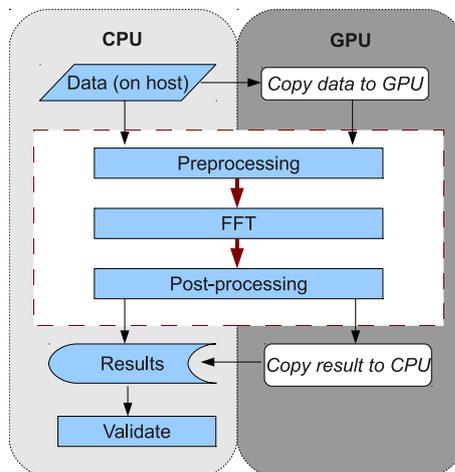


Figure 9.3: The flowchart for validating GPU results.

The work described in the rest of this chapter, evaluates the performance of GPU based processing and post-processing of OCT data. GPU based processing is benchmarked against CPU processing speeds. The processing speed and

numerical accuracy of the results for different GPUs (NVIDIA Tesla C1069, GT240, and GeForce GTX) are evaluated.

9.3 GPU implementation

Four different NVIDIA GPUs¹ (Tesla C1060, GeForce GT 240, GeForce 9400M and Quadro 6000) were used in this study, whilst a Tesla C1060 was used as the primary testing platform. The specifications of the GPUs are summarised in Table 9.1, and detailed specifications can be found elsewhere [107].

The Tesla C1060 and GeForce GT 240 GPUs are hosted in a PC workstation (Intel Core i7, 2.67 GHz, 6 GB DRAM) running the Ubuntu 9.10 Linux operating system. A MacBook running MacOS Lion (Intel Core 2 Duo, 2 GHz, 2 GB DRAM) hosts the integrated GPU GeForce 9400M. The latest GPU based on NVIDIA Fermi architecture, Quadro 6000 is hosted by a PC running Window 7 operating system (Dell Precision T7500, Intel Xeon E5620 @2.4Hz, 12 GB DRAM).

The development work started in CUDA version 3.0 and later migrated to version 4.0. The custom programs were written in C using NVIDIA's CUDA programming extensions and executed with the CUDA driver and tools. A cross platform build system (CMake version 2.8.6) was used for building and packaging the software. CMake allows one to compile the code independent of the development environment, generating native Makefiles using compiler independent configurations.

In Ubuntu and MacOS a GNU compiler (version 4.4.1) was used, whilst in Windows 7 native Microsoft C/C++ compilers and linkers were used for compiling and linking along with NVIDIA CUDA compiler (NVCC versions 3.2 and above). NVCC splits the CUDA source files into GPU and CPU specific codes. GPU codes are compiled to parallel thread execution (PTX)² (parallel thread execution) by the CUDA compiler whilst the C/C++ portion of the code was compiled using native C/C++ compilers.

²Intermediate assembler language used by NVIDIA Compilers

Table 9.1: *Details of the GPUs used in this study*

	Tesla C1060	GeForce GT 240
Global memory	4 GBytes	532 MBytes
Multi Processors	30	12
Cores (Streaming Processors)	240	96
Total Constant Memory (Bytes)	16 384	16 384
Total Shared Memory (Bytes)	16 384	16 384
Total Registers (Bytes)	16 384	16 384
Maximum threads per block	756	512
Maximum active threads per block	1024	1024
Theoretical processing (GFlops)	933 (72)	385.9
Clock Rate (GHz)	1.30 GHz	1.32 GHz

The GPUs have different memory spaces which must be correctly exploited to achieve maximum performance. In order to increase the performance all the constant variables required for the computations are copied to the GPU's constant memory. Lagrangian interpolation coefficients are computed on the CPU copied to the GPU constant memory along with the background and the resampling table during the GPU initialisation stage. Figure 9.4 illustrates the transferring of constant data to GPU constant memory from the CPU memory.

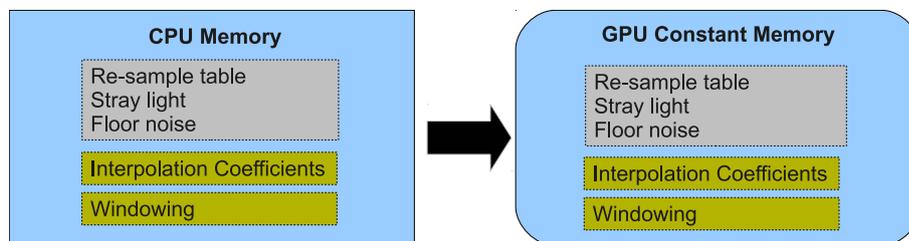


Figure 9.4: *Copy data from CPU memory to GPU Constant memory at the initialisation state. Interpolation coefficients are precomputed on CPU and then transferred to GPU constant memory*

The standard OCT data processing in GPU is split into three kernels:

- **Pre-process kernel** which subtracts the background and interpolates the data using either linear or Lagrangian interpolation and finally multiplies the interpolated data by a windowing function. Figure 9.5 illustrates the flowchart of pre-processing steps. Data in different blocks cannot communicate and there is no dependency in the block execution by default. Therefore

the algorithm is designed so each A-scan is processed by a single block.

- **FFT kernel** which takes the modulus of the complex FFT of the resampled data, using the CUFFT library in the GPU.
- **Post-process kernel** which computes the magnitude of the complex values and takes the logarithm (dB scale). Figure 9.6 illustrates the flowchart of the post-processing steps.

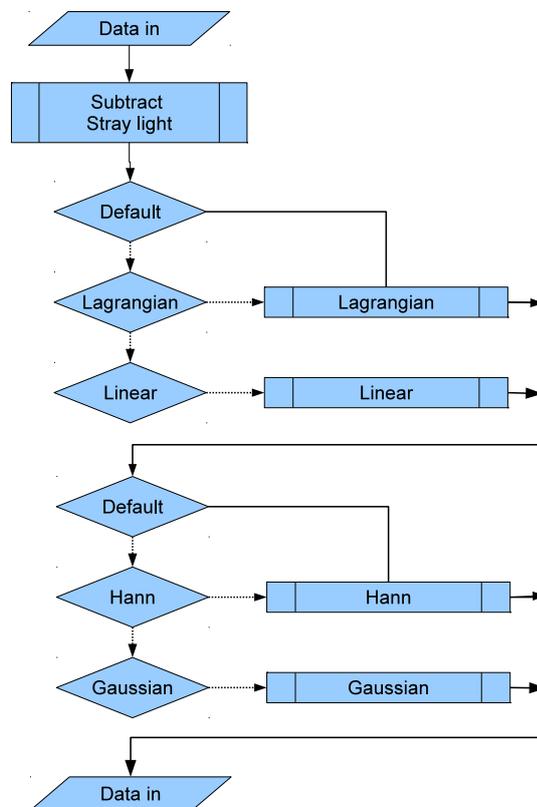


Figure 9.5: Data flowchart during the pre-processing. Default paths are represented by lines and the dot-dot lines represent the alternative paths taken by the kernel, depending on the execution conditions.

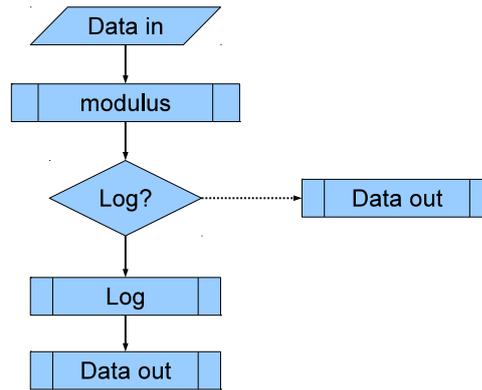


Figure 9.6: Flow chart of post-processing, where dot-dot lines represent the alternative paths.

9.3.1 The kernels configuration

Kernels require configuration parameters at runtime, to configure the threads and blocks. For example a kernel can be invoked by CPU as follows:

```
1 preProcess$<<<$NumBlocks, NumThreads>>>(Args) ;
```

where `preProcess` is the name of the kernel, $NumBlocks(N_{Blocks})$ and $NumThreads(N_{Threads})$ are the number of blocks per grid and the number of threads per block allocated. Total number of threads allocated $T = N_{Blocks}N_{Threads}$ must be greater than the number of elements in the data. Arrangement of the threads per block and blocks per grid is arbitrary but can greatly influence the processing rate of the kernel. The number of threads allocated per block can be arbitrary but must be between $2 \times \text{warp size}$ and smaller than the maximum threads per block (i.e 512). In order to achieve maximum performance, the threads must be multiples of the warp size (i.e. $32n$ where $n \in [2, 16]$).

Figure 9.7 shows an example grid with 15 blocks, where each block consists of 64 threads. Blocks are executed independently and therefore threads within different blocks cannot communicate with each other.

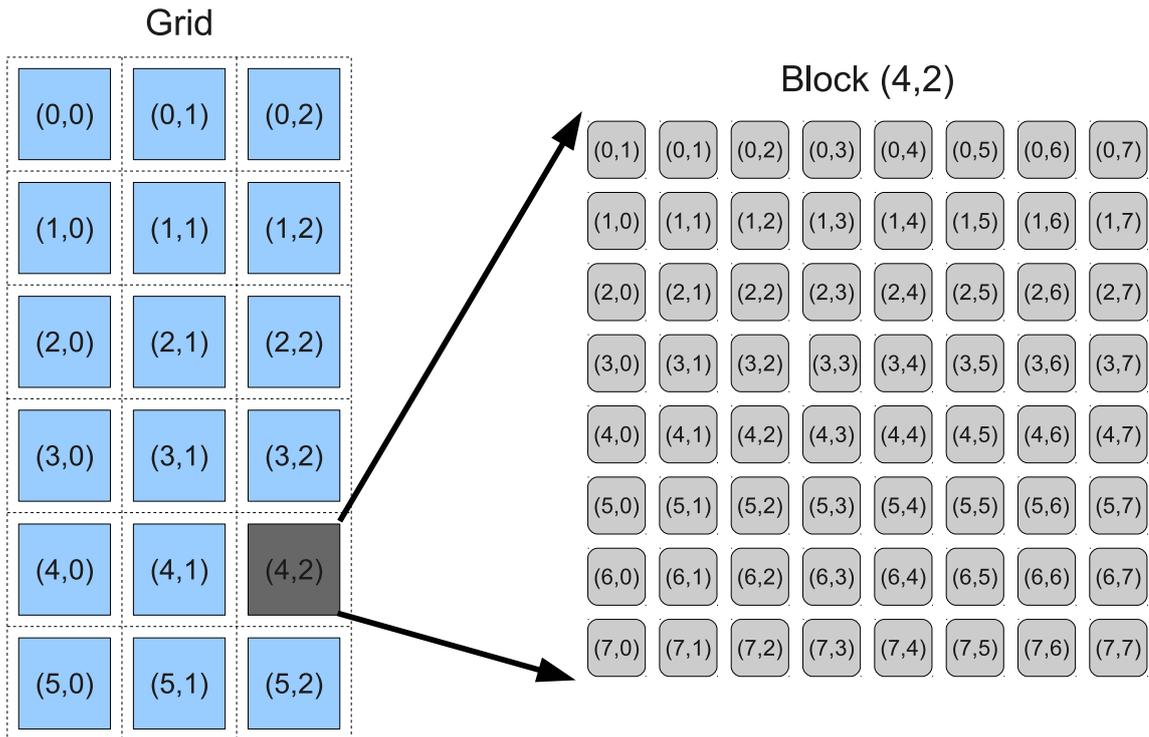


Figure 9.7: Example kernel configuration showing a 3×5 blocks arranged in a 2D grid, where each grid is composed of 8×8 threads

The number of threads per block that gives maximum performance can be estimated by:

$$\mathbb{T} = \text{mod}(N_B, N_{SM}/2), \quad (9.3.1)$$

where N_B and N_{SM} are the number of blocks required and the number of SMs available in the GPU. When \mathbb{T} is close to zero the processing rate is expected to be higher. This is to do with the workload sharing between the SMs. For example, let us consider a GPU with 12 SMs used for processing 7680 elements. Allocating 512 threads per block requires 15 blocks to process the data. Schedule assigns two blocks to three of the SMs. These three blocks will execute the second block whilst the other blocks stays idle. However, if 128 threads are allocated per block, a total of 60 blocks are required to process the data. Then the workload is equally distributed between all the SMs (i.e. 5 blocks per SM) thus fewer processors go to an idle state and thus it can be expected that the performance improves.

pre-processing kernel

Processing an A-scan of 1024 elements (where each element is 16 bits or 2 Bytes) requires 2 kB of shared memory per block regardless of the interpolation routine used. Each SM has 16 kB of shared memory, therefore a maximum of 8 blocks can be simultaneously active on each SM provided there are enough other resources such as registers available (i.e. shared memory limits the number of active blocks per SM to eight). A minimum of two active warps (64 threads) per block are required to hide the long latency process by means of warp swapping.

Each element in the A-scan is mapped to a thread within the block containing the T threads such that:

$$f : d \mapsto \text{mod}(d, T) \quad (9.3.2)$$

$$d \in 0 : A - 1 \quad (9.3.3)$$

where A is the size of the A-scan and d is the index of the data. Each block is mapped to one of the SM in the following manner:

$$f : b \mapsto \text{mod}(b, N_{SM}) \quad (9.3.4)$$

$$b \in 0 : B - 1 \quad (9.3.5)$$

where N_{SM} is the number of SMs available in the GPU, B is the number of blocks required for processing the data and b is the index of the block.

Post processing kernel

For an A-scan of length L , the real to complex CUFFT produces $(L/2 + 1)$ elements of complex numbers. Absolute values (modulus) of these are computed by the post-processing kernel. There is no dependency between the elements so the kernel configuration is much simpler. The total number of complex elements

E is given by:

$$E = \left(\frac{L}{2} + 1 \right) N_A \quad (9.3.6)$$

where N_A is the number of A-scans. The number of blocks with T threads required to process the data with E size is:

$$B = \left\lceil \frac{E}{T} \right\rceil. \quad (9.3.7)$$

9.4 Methods

9.4.1 Processing rate

Accessing the GPU memory blocks the CPU, but the kernel executions are asynchronous so the handle is returned to the CPU after invoking each kernel. Explicit synchronisation is necessary to block the handle returning to the CPU. Figure 9.8 illustrates the method of timing data transfer and kernel execution. CUDA event handles were used for recording the time before and at the end of the data transfer and kernel execution. The elapsed time Δt is calculated as:

$$\Delta t = t_e - t_s, \quad (9.4.1)$$

where t_s and t_e are the time recorded at the start and end of the process in by CUDA event handle. The processing rate and copying rate R (number of lines per second) is calculated as:

$$R = \frac{N_A}{\Delta t}, \quad (9.4.2)$$

where N_A is the number of A-scans processed.

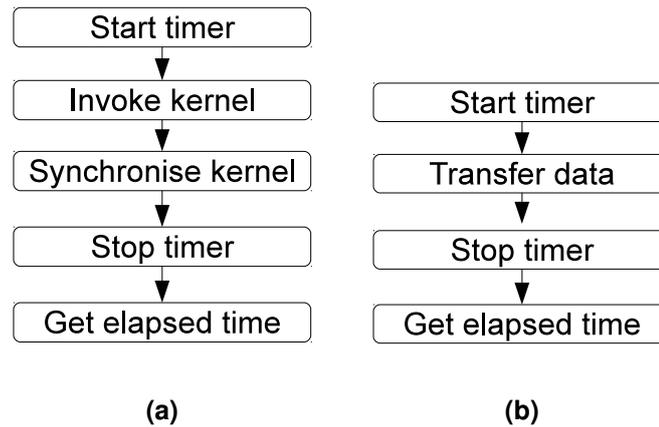


Figure 9.8: Flowchart for timing a) CUDA kernels and b) data transfer.

By default, the GPU memory access is blocked during kernel execution. However, for concurrent copy and execution, the data is divided into multiple streams. The asynchronous memory copy function allows the transferring of data to and from the GPU while a kernel is executed.

Most of the current GPUs allow concurrent copy and execution. The data is split into different streams (sub-sets). The copying and executing of the streams are independent of each other. The GPU can copy a stream of data while processing another one. In this way only the first stream requires additional copying time. The processing of the streams on the GPU hides the copying time for the rest of the streams. Therefore the copying time of a stream should be smaller than and proximate to the processing time of streams.

Figure 9.9 illustrates sequential processing and two types of asynchronous processing (i.e. concurrent copy and executions). Where H2D and D2H refer to the data transfer from 'Host to Device' and 'Device to Host' and kernel 1 and 2 refer to two kernels that process the data in a serial manner. The kernels could be executed in two ways as shown in Figure 9.9 b and c. The latter is preferable and a more convenient approach, that is implemented in our routines, where the total execution time is useful.

CHAPTER 9. DATA PROCESSING

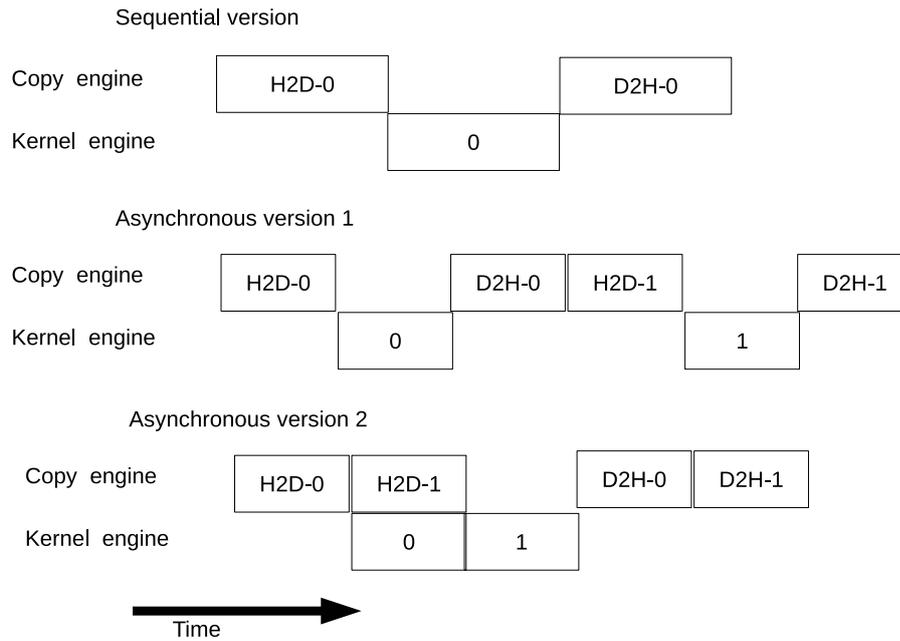


Figure 9.9: *Tesla C1060 execution time line.*

9.5 Results and Discussion

Figure 9.10 shows the pre-processing rate for a fixed number of A-scans, with a varying number of blocks, for the *Tesla C1060* and *GeForce GT 240* GPUs. The pre-processing rate increases as the number of blocks reaches the number of SMs in the GPU. To keep the GPU busy, the number of blocks should be greater than the number of SMs. Pre-processing implementation achieves the maximum rate when the number of blocks equals the number of SMs. The post processing achieves the maximum rate when the number of blocks is much more than than the available number of SMs.

The processing time taken by each kernel (pre process, FFT and post process), as well as pre-processing employing two different interpolations spline and Lagrange interpolation, were measured. Figure 9.11 shows the overall processing rate for 10,000 A-scans on *Tesla C1060* and *GeForce GT 240* GPUs recorded for 140 runs. Table 9.2 illustrates the processing values for three GPUs. It shows that the processing rate achievable by *Tesla C1060* is an order of mag-

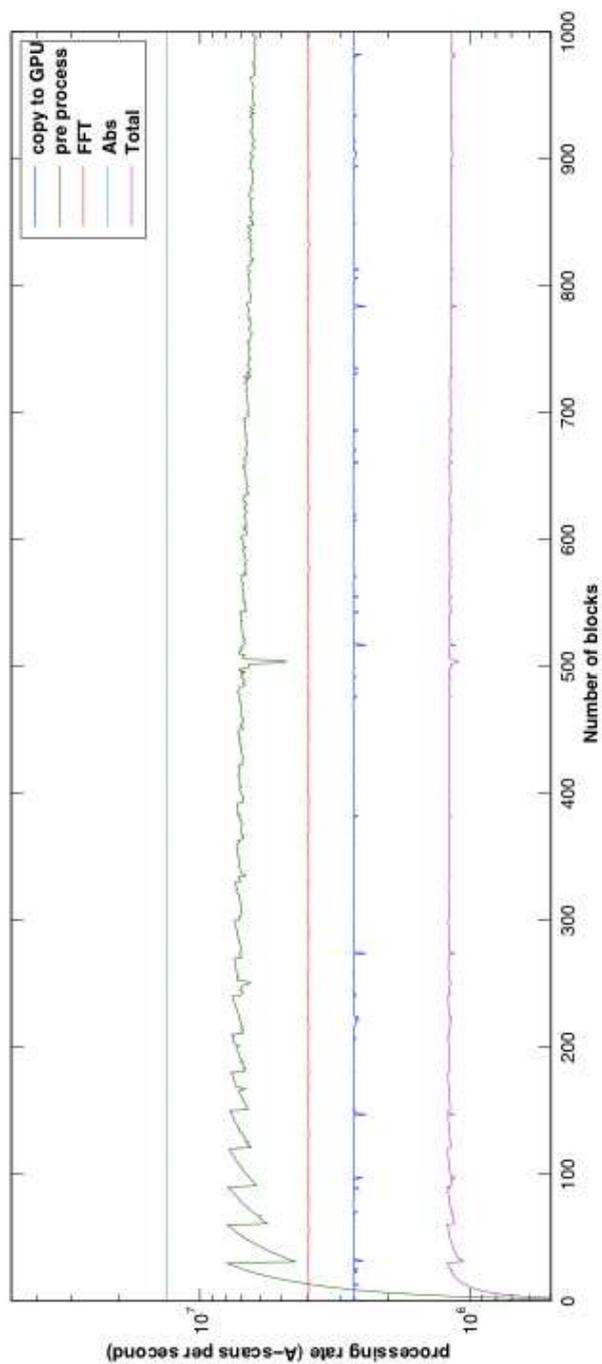


Figure 9.10: Maximum and standard deviation of the processing speed achieved with concurrent copy and transfer method for 100,000 A-scans.

CHAPTER 9. DATA PROCESSING

nitude greater than the *GeForce GT 240* and two times faster than *GeForce 9000M* GPUs. This is an expected behaviour, since *Tesla C1060* has higher number of processors (i.e GFlops) than the other two GPUs. The theoretical performance of the *Tesla C1060* is about 2.4 times higher than that of the *GeForce GT 240*.

Table 9.2: Max, Min, mean and standard deviation of the processing rates for three GPUs

Hardware	Max	Min	Mean	Std. deviation
Tesla	1.2768×10^6	1.2543×10^6	1.2641×10^6	3.7593×10^3
GT240	1.2727×10^5	7.5347×10^4	9.7523×10^4	2.2889×10^4
GeForce	6.0115×10^5	5.9572×10^5	5.9859×10^5	9.9469×10^2

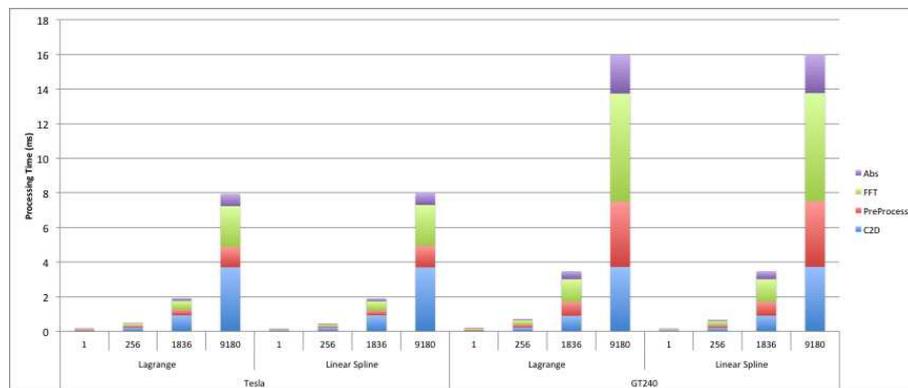


Figure 9.12: The processing time of different kernels measured on Tesla 1060 and GT240 GPUs. The processing time was recorded for both the Lagrange and the linear-spline interpolations for varying number of A-scans.

For a lower number of A-scans, the higher percentage of the processing time is spent on computations rather than on data transfer. This is due to the fact that GPU is not fully utilised at lower A-scans and the amount of data transferred is much smaller than the bandwidth, thus the full processing potential is not realised. However this situation is reversed when large amounts of data is processed. The data transfer becomes the bottleneck, since the data transfer time is greater than processing time. This is attributed to the memory bandwidth of the GPUs. In this case the full processing potential of the GPU is realised. Both GPU have the same memory bandwidth, thus the data transfer takes almost the same amount of time, as shown in Figure 9.12. However, the number of processors available

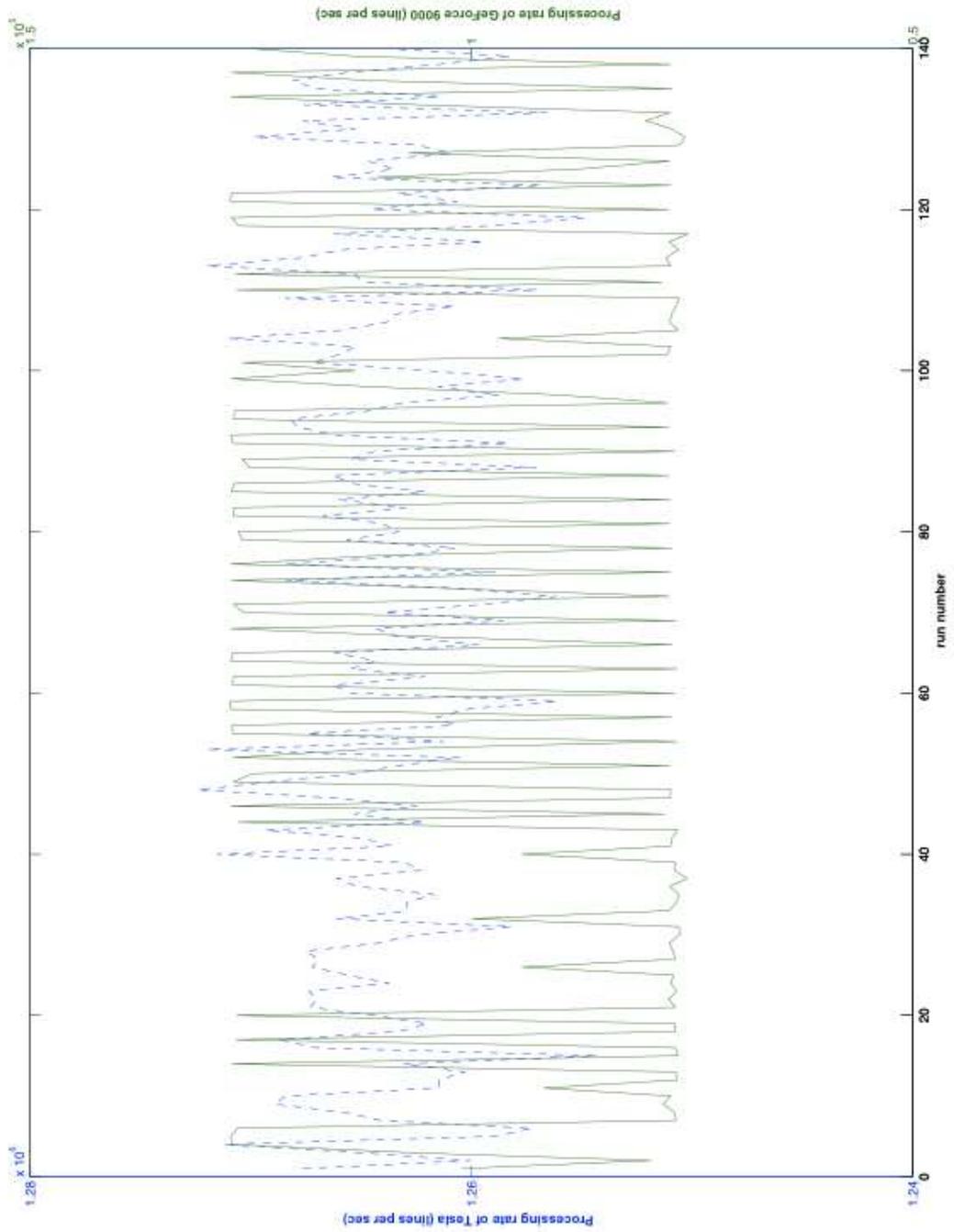


Figure 9.11: Processing rate for 140 different runs for Tesla C1060 and GeForce 9000M

CHAPTER 9. DATA PROCESSING

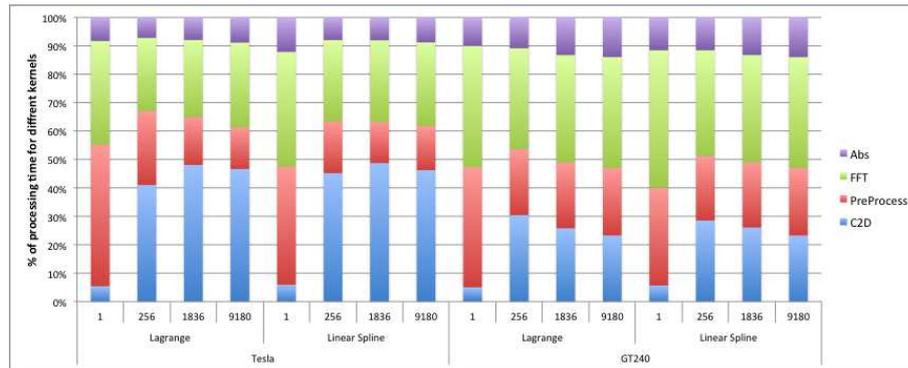


Figure 9.13: Processing time taken for different tasks (kernel) as percentage of total processing time on different GPU and for different interpolations.

for the computation (i.e. GFlops) differs, which results in variation amongst the observed processing rates between the GPUs.

Figure 9.13 shows the processing time as a percentage of overall time for different number of A-scans on the *Tesla C1060* and *GeForce GT 240* GPUs. It is clear that, although the copying time is the same for both GPUs, the percentage of the overall time spent on copying differs.

By exploiting the concurrent copy and execution method, the copying of the data can be hidden. In this method the data is divided into multiple streams, each is copied and processed independently. Concurrent copy and execution could be employed to remove this data transfer bottleneck. The data is divided into multiple sub-sets, and computations/execution of one set is overlapped with copying of another. In this case only the first section of the data is required for dedicated copying time, all other sub-volumes are overlapped with execution, which dramatically improves the results.

Table 9.3 shows the processing rate achieved by the pre-processing kernel, without concurrent copy and transfer, for varying number of threads per blocks along with the expected number of active blocks per SM for pre-processing and post-processing kernels.

Table 9.3: Average pre-processing rate on Tesla for varying number of threads per block.

Number of threads	Processing rate (lines per sec)		Active blocks
	Tesla C1060	Tesla C1060	
128	994391.625	1030414.125	8
256	1279076.125	1044665.250	4
512	1269830.375	1015378.875	2

Figure 9.14 shows the processing rate of different processes and the overall rate for varying number of A-scans, without concurrent copy and execution. Bash shell script is used for invoking the programme multiple times, and the processing rate is averaged over 100 runs. The maximum processing rate of $1.2641 \times 10^6 \pm 3.7593 \times 10^3 Hz$ is reached at nearly 3000 A-scans. The observed fluctuations of the processing rate may be caused by the context swapping in GPUs.

Furthermore, a processing rate of over 410,000 A-scans per second is achieved with a lower end (GeForce GT 240) GPU. This is about 45 times faster than a single threaded CPU based processing. For most of the Fourier-Domain OCT system, the data acquisition become the new bottleneck. Therefore without any additional hardware, real time processing can be achieved for most of the OCT system. However, the advantage of using a dedicated GPU is that it provided sufficient additional computational capability for more complex analysis and data processing to be accomplished in real-time.

Taking the advantage of concurrent copy and execution methods the processing time can be further enhanced. Figure 9.15 shows the times for processing the first stream and the total processing time for varying number of A-scans. The number of A-scans per stream is varied. A maximum processing of over 2.125866 million A-scans per second is reached on Tesla when the number of A-scans in the stream is about 4000.

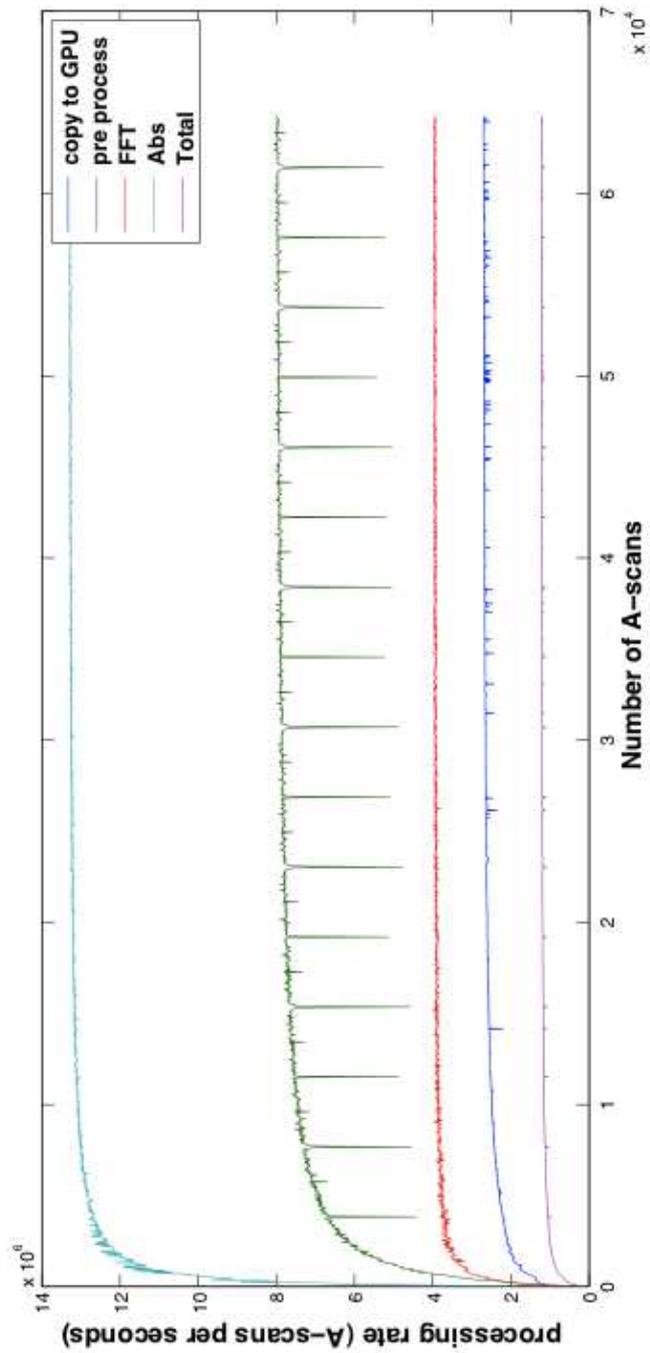


Figure 9.14: Processing speed with varying number of A-scans in batch, for Tesla GPU.

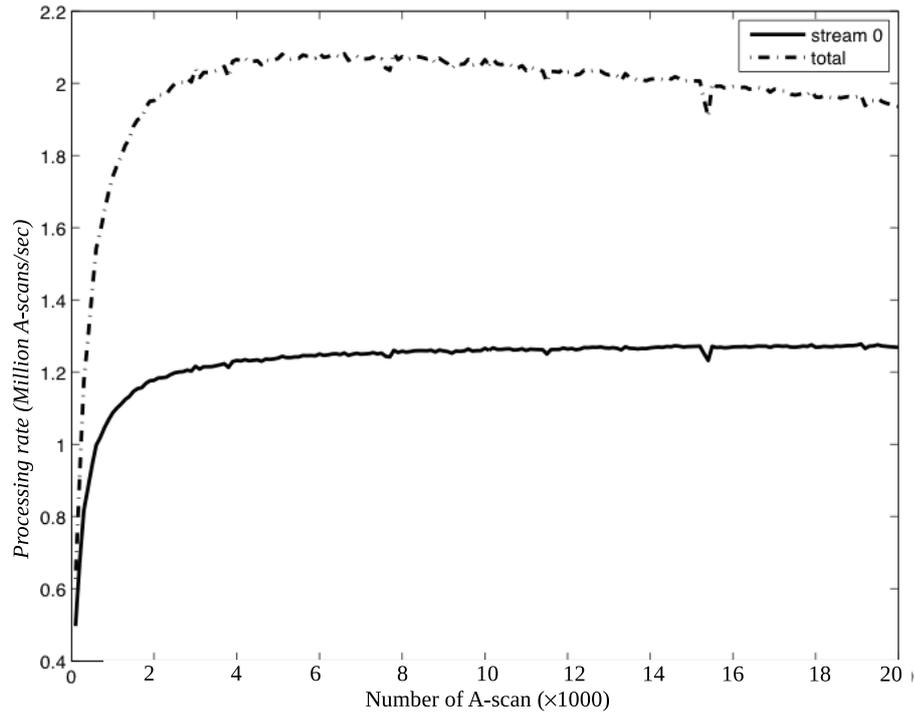


Figure 9.15: Processing rate achieved, using concurrent copy and execution, for varying number of A-scans. stream-0 is the first stream. Total is the total processing time.

9.6 Conclusion

A maximum processing rate of over $2.125866 \times 10^6 \pm 3.7593 \times 10^3 \text{ Hz}$ and $5.9859 \times 10^5 \pm 9.9469 \times 10^2 \text{ Hz}$ are achieved on the Tesla and GT240 GPUs respectively. Additionally, nearly $9.7523 \times 10^4 \pm 2.2880 \times 10^2 \text{ Hz}$ A-scans are achieved, without concurrent copy and transfer, on a GeForce GPU which is one of the lowest end GPUs. The processing capability of the dedicated GPUs are 2.836 orders of magnitude greater than the current FD-OCT acquisition systems. This means that more complex post processing techniques such as filters, noise removal etc. can be implemented in real time.

Advanced GPU based post-processing methods such as deconvolution can be used to enhance the quality of the OCT image in real time. Implementation of GPU based deconvolution is described at the end of this chapter.

CHAPTER 9. DATA PROCESSING

Theoretical (single precision) floating point performance of Quadro 6000 GPU is 1030.4 Gigaflops, which is greater than that of *Tesla C1060* . Although no benchmarking studies have been conducted on the Quadro GPU, we can expect an increased processing rates.

Chapter 10

Real Time Optical Coherence

Tomography

10.1 Introduction

This chapter briefly describes the development of an SB-OCT system (called here the RT-OCT system) that will acquire, process and render the results in real time. This system was developed for use in clinical trials in dentistry. Since we have already developed the GPU based processing that achieves the required processing needs to render results in real time, Dr Peter H Tomlins at Queen Mary University invited me to work on this project with him. The objectives of this project were to set-up the optical system (i.e. the interferometer), construct a high speed spectrometer, integrate GPU based processing and rendering routines and develop a graphical user interface (GUI) to acquire, process and display the data.

My contributions to this work are the following: I set up the initial optical layout based on P.H.Tomlins drawings, wrote the initial code to connect the DAQ to the camera to generate trigger signals and acquire data, whilst P.Tomlins has contributed to the generation of the signal for driving galvo mirrors. I have also implemented the GPU based processing to process the acquired data. Later these implementation were developed and enhanced further by P.H. Tomlins. He also

wrote the GUI of this software, for displaying the results and interacting with the system.

10.1.1 Description of the OCT system

The schematic illustration of the RT-OCT system is shown in Figure 10.1. The optical layout is based around a Michelson interferometer arrangement. Two fibre optical splitters (OS_1 and OS_2) with a 90 : 10 coupling ratio and two optical circulator (OC_1 and OC_2) were connected as shown.

A super-luminescent diode (SLD1325, Thorlabs) that emits light at a centre wavelength of 1325 nm with a FWHM spectral width greater than 100 nm was used as the light source. This SLED is mounted in a 14-pin butterfly laser diode mount (LM14S2, Thorlabs), and the operating properties of the diode were controlled using an SLED Driver (ITC4001). This diode could give a theoretical axial resolution of approximately $7.7\text{ }\mu\text{m}$ measured in air.

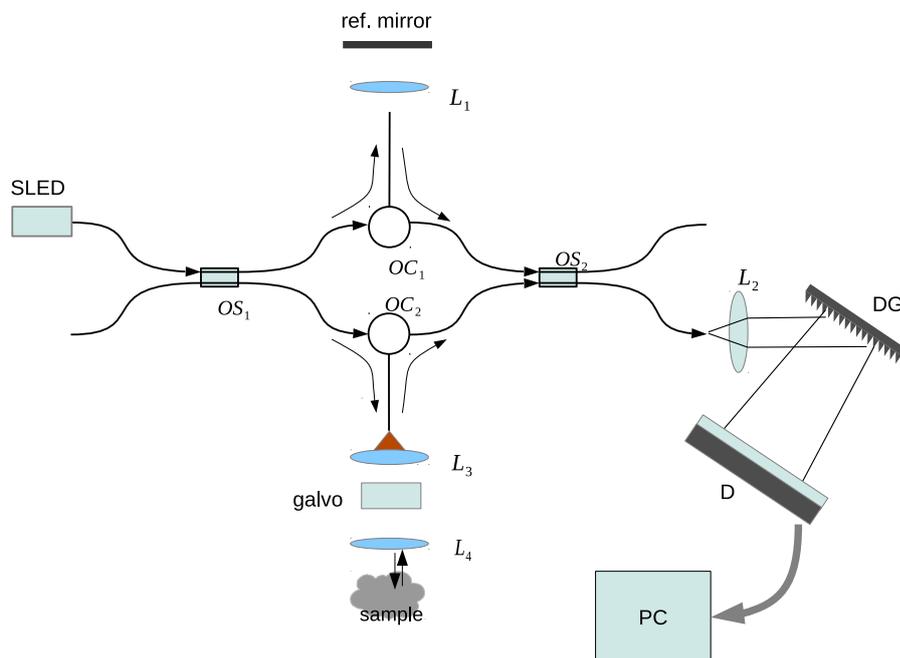


Figure 10.1: Optical layout of the RT-OCT. SLED, super-luminescent diode; L, achromatic lens; DG, diffraction grating, OC, optical circulator, OS optical splitter, D Detector (CMOS line scan camera), PC personal computer.

The output of the broadband SLED is coupled into one arm (r) of the first optical splitter (OS_1), at the exit of the OS_1 light is split along the sample and reference arms. Light that travels along the reference arm, enters first arm of the optical circulator (OC_1), is reflected by the reference mirror (with phase delay) and exits through the third arm of the OC_1 . Light that travels along the sample arms, enters first arm of the OC_2 , is reflected by the refractive index discontinuities within the sample. The combined light from all layers travels through the through the third arm of the OC_2 . Light from both optical circulators are combined at the entrance of the second optical splitter OS_2 .

The sample arm has a two dimensional (XY) galvo scanner for scanning the sample. The galvo scanner is driven by signal generated by the DAQ (PCIe6323, National Instruments, Berkshire). The light comes out of the sample arm and is collimated using lens 3 (L_3). The collimated beam is first reflected by the mirror on the galvoX and then redirected to the mirror on galvoY. Light reflected from mirror on galvoY is focused by another lens (L_4 coupled into object lens optical relay line) into the sample.

The sample arm consists of a collimator, pair of galvanometer mirrors for transverse scanning (orthogonal to each other) and a focusing lens, as shown in Figure 10.1. The collimated beam is redirected to different spatial locations within the scanning/image domain by the scanning mirrors, where it is focused into the sample by the focusing lens. Galvanometer mirrors are continuously scanned using sawtooth profiles, whilst the camera is configured to obtain the A-scans at equal intervals.

Back reflected signals from the reference mirror and sample are recombined at the second coupler. At the exit the signal is split again with the ratio of 1 : 9. The recombined light returning from the sample and reference arms is detected using a spectrometer set-up. The light is collimated using an achromatic lens, L, then dispersed into difference components of wavelength by a reflective grating and then focused onto the detector using another achromatic lens. The detector is an Indium gallium arsenide (InGaAs) line scan camera (Goodrich, Princeton,

NJ) operating at 92 kHz (i.e. lines per sec) consisting of 1024 pixels. Each pixel is $14\mu\text{m} \times 14\mu\text{m}$ in size and 14 bits in depth. The camera has 1024 pixels, width $24\mu\text{m}$ width and 12 bit data resolution, and sensitive to the wavelengths between 0.8 to $1.7\ \mu\text{m}$.

A DAQ board, installed on one of the Peripheral Component Interconnect Express (PCIe) slot, was used to generate trigger signals and the drive voltage (signal) for driving galvanometer mirrors. The DAQ was connected to a connector block (SCB-68, National Instruments, Berkshire) via a cable (SHC6D-88 EPM). This DAQ has 32 (16 differential) analogue input and four analogue outputs, both with 16 *bit* resolution and a voltage range between -10 V and 10 V . It has 48 digital I/O with a frequency of 1 MHz that is comparable of generating/receiving signal in the range of $[0, 5\text{ V}]$, and has 32 *bit* resolution and 100 MHz source frequency. The connector block used has half the number of analogue and digital interfaces that are available on each DAQ board.

The scanners have a maximum scan angle (mechanical) of $\pm 12.5\text{ deg}$ for an input voltage of $\pm 10\text{ V}$. This corresponds to 0.8 V being converted to one degree of mechanical rotation where the relationship between voltage and rotation is linear. The mirrors can oscillate with a maximum frequency of 100 Hz and 350 Hz when operated with square wave and sine wave mode respectively. During this work, a sawtooth profile was generated with 100 Hz and 1 Hz for the fast and slow axis respectively (this give a total scanning time for a sample with 1000×1000 A-scans as $< 11\text{ s}$ (i.e. 92 frames or $92k\text{ A} - \text{scans}/\text{sec}$). The speed here is limited by the camera since it can only record up to $92k\text{ A} - \text{scans}/\text{sec}$. We have also noticed, that there is a reduction of acquisition speed caused by reading/writing to and from the memory allocated by National Instruments (NI) library. This need to be solved, in future, to attain the maximum speed of the camera.

For a fixed incident beam and rotating mirror, the angle of reflected light is twice the angle of rotation. (i.e. one degree mechanical rotation of the mirror will case two degree change in the beam divergence). This is illustrated in Figure 10.2. The galvanometer used in this study has $\pm 12.5\text{ deg}$ mechanical rotation

CHAPTER 10. REAL TIME OPTICAL COHERENCE TOMOGRAPHY

that corresponding to ± 25 deg optical rotation. The maximum lateral scan L is related to the angle of reflection by the galvanometer mirror θ and the focal distance between the lens $L1$.

$$L = H \tan(\theta) \quad (10.1.1)$$

where H is focal length of L1. For a typical H value of 100 mm , this gives a scan range of 46.63 mm .

A pair of analogue outputs was used for generating the galvanometer driving signals. The DAQ card used for generating the driver signal has a digital-to-analogue converter (DAC) resolution of 16 bit . Hence, for the voltage range of 20 V , this card can generate a voltage that results in a resolution of $305.175 \mu\text{V}$ obtained as follows:

$$\begin{aligned} \text{Resolution} &= \frac{\text{voltage range}}{\text{DAC resolution}} & (10.1.2) \\ &= 20/2^{16} \\ &= 305.175 \mu\text{V}. \end{aligned}$$

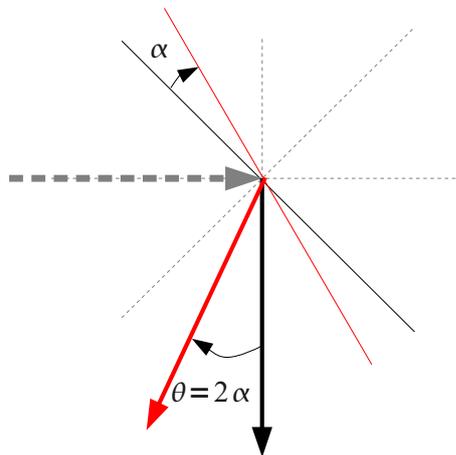


Figure 10.2: Mirror rotation and optical scan range

Using the source frequency counter as the base time, digital output signals were generated to trigger each A-scan, and controlling the integration time. The

camera is triggered using an external trigger (A-scans), which is provided by the DAQ. External triggering of the camera (for each A-scan) ensures that there is no overlap between the A-scan capture and galvanometer movement. They were all synchronised using the source frequency clock available in the DAQ board.

Data from the camera were captured using the frame grabber (PCIe 1427, National Instruments, Berkshire) connected via a camera link cable. The frame grabber has a maximum capture rate of 100 kHz , that is slightly higher than the rate of the camera. External triggers, produced by DAQ, were used for triggering the capture process. This has no internal memory, but with high bandwidth (i.e. 200 MB/s via a PCIe $1\times$) data can be streamed to the PC memory (DRAM).

10.1.2 Description of the PC/software model

A 64 bit PC (Dell Precision T7500) system with Intel Xeon Quad core CPU (E5260, 2.4 Hz) having DRAM¹ of 12 GB running Microsoft's Window 7 OS was used for [controlling] the OCT. This PC system consists of a dedicated GPU (NVIDIA, Quadro 600), a frame grabber and a DAQ were installed on using PCIe express slots. The GPU is used for display as well as accelerating the OCT data process, and frame grabber and DAQ were used for acquiring the data from the camera using CameraLink cable and generate signals/triggers that drive the galvo and trigger the data acquisition.

Microsoft Visual C++ Studio 2010 was used as the development environment, and configured to use the CUDA run-time libraries, compiler (NVIDIA CUDA compiler (NVCC)), FFTW library, and NI Libraries. A cross platform build system, CMAKE (version 2.8.6), was used to build the software. A compiler independent configuration of CMAKE helped in porting the software developed under Unix environment to the Microsoft environment. CMAKE generates native make file that use native compilers for compilation of the software. The CPU part of the code is compiled using Microsoft compiler and GPU part of the codes is compiled using

¹double data rate type three synchronous dynamic random access memory (DDR3 SDRAM)

CUDA compilers.

Function wrappers written in C/C++ around the NI DLL and CUDA processing DLL were used for the interaction with the frame grabber and the data processing using the GPU. In the current configuration, the NI allocates its own memory and stores the captured data in it. However, using low level code we could take control of this. Using external triggers produced by the DAQ, A-scans were captured and stored in the NI memory space. When the entire volumetric data is obtained, the data is transferred from the NI memory space to the CUDA CPU memory space. Then the data is processed using the GPU based processing method already described in Section 9. In addition to the standard processing, additional computations are included such as: 1) en-face image 2)attenuation depth images, were also computed on GPU. Processed data were then copied back to CPU and rendered using the NI rendering libraries.

Figure 10.3 shows the morphological visualisation of mucocele, and Figure 10.4 shows images obtained from one of the PSF phantom (polyurethane matrix filled with a low density of iron oxide). These images were obtained by Peter. H. Tomlins.

Using fs inscribed calibration phantom we have noticed distortion at the edges of the B-scan. Artefacts and noise can arise from mechanical and optical issues such as scanning errors and dispersion mismatch. This project is in its early phase and future development will look at the way to improve the detected signal and also quantitatively verify the performance of this system using our phantoms.

10.2 Conclusion

This chapter describes the culmination of work in this thesis where the real time processing and artefacts were brought together to improve and evaluate the performance of an OCT system. A SB-OCT system was developed, that is capable of producing real time images. The processing of the OCT data were accelerated using GPU based algorithms. Work is still in its early phase. The system was

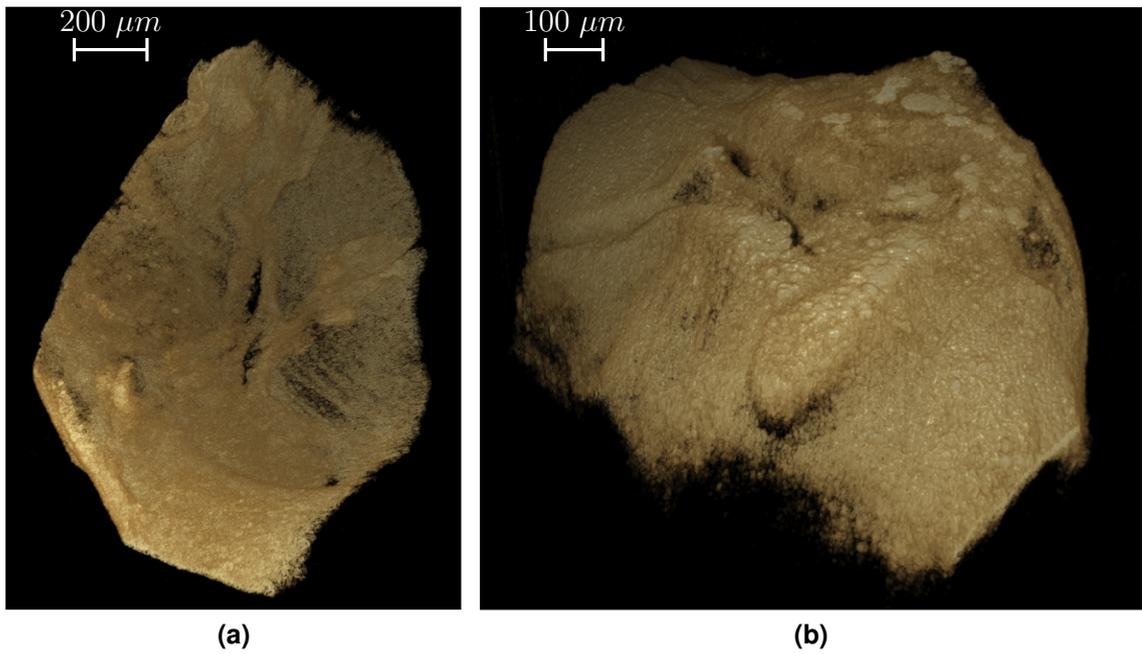


Figure 10.3: *The morphological visualisation of mucocele obtained using RT-OCT. a) Top view and b) internal view.*

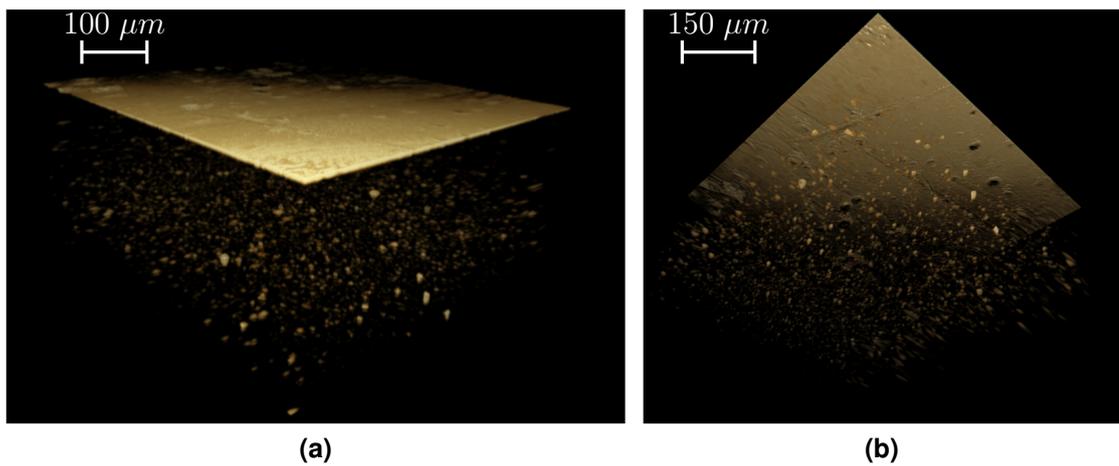


Figure 10.4: *Images of the PSF phantom obtained by RT-OCT.*

CHAPTER 10. REAL TIME OPTICAL COHERENCE TOMOGRAPHY

characterised using the OCT-phantoms developed previously.

Chapter 11

Conclusion

11.1 Conclusion

This thesis investigates the possibility of fabricating phantoms using fs laser inscription that can be adopted as standards for calibration of any OCT system.

In addition, it has also been demonstrated the potential of GPU based data processing methods to accelerate the processing speed of OCT data and overcome a significant bottleneck in SD-OCT systems. Other post-processing methods have also been implemented. It has been shown that GPU processing offers the potential of applying advance post processing methods to OCT data in real time.

The motivation for the work presented in this thesis and the background details of the three key areas (OCT, fs inscription and GPU processing) have been covered in Chapters 1 and 2. The remainder of the thesis describes the novel work carried out including the investigation and optimization of the fs inscription parameters for OCT phantom fabrication, the design and characterization of phantoms, the development of quantitative methods and GPU based processing. The current status of the work and the planned potential future development are given in the first and the second section of this chapter respectively.

The requirements for ideal phantoms and the feasibility of manufacturing using

CHAPTER 11. CONCLUSION

fs laser inscription have been investigated. In order to measure the resolution of the system directly, we need to produce structures that are smaller than the resolution of the system being studied. The effects of a varying structure size on the measured FWHM width has been investigated using simulation, and it was concluded that structures of around $1 \mu m$ size are required for current OCT systems. The size of the structures produced by the fs laser is related to the confocal parameter and beam width at focus. It is possible to fabricate structures with $1 \mu m$ in size using high NA objectives.

In order to evaluate the suitable inscription parameters, we have conducted parameter space test studies using three fs laser systems that have different operating parameters. Structures with different energy, speed and depth were inscribed in a variety of transparent material including fused silica, borosilicate, BK7, PMMA and polyurethane resin. The produced structures were characterized using an optical microscope equipped with QPM and several OCT systems. The cross sectional morphology of some of the initial samples was studied by cleaving and chemically etching the cleaved side.

The results show that the size of the structure inscribed decreases with increasing depth (at constant energy and speed), decreases with decreasing energy (at constant depth and speed), and decreases with increasing speed. This is due to the overall dependence on fluence.

Inscription with the *Low-Rep 800 nm* system (repetition rate $1 kHz$) produced very periodic pearl chain like structures when the sample was translated at a speed of $1 mm/s$. At slightly lower speed, $0.8 - 1 mm/s$ it also produced reproducible structures. The phase of the inscribed region increases with increasing energy, but with decreasing speed it becomes negative. This indicates material damage rather than refractive index change. The sizes of the structures were limited by the focusing objective.

Inscription with the *Low-Rep 1026 nm* system (at $100 kHz$) produced optimal damage (i.e. finer edges than at higher speed) at the translation speed of $1 mm/s$. The size of the structures (lateral width) increases with increasing energy, and

CHAPTER 11. CONCLUSION

decreasing speed. The size of the induced structures decreases with increasing depth. This system produces around $1.7 \mu\text{m}$ structures, that are smaller than the required size for OCT-phantom. However the inscribed structures are elongated, this could be reduced by using additional optics (i.e. slits or cylindrical lens) to reduce the elongation. We have done some preliminary investigations on this but a full parameter space test is yet to be done.

The *High-Rep 800 nm* system produced optical fibre like structures, that on first inspection would be ideal for our work since optically they appear to be small and of high quality (i.e. specifically for sensitivity test). Unfortunately, the inscribed structures were not visible in the B-scan of MD-OCT, Santec-OCT or Thorlab OCT systems. However, *en-face* images obtained by an RT-OCT system shows inscribed regions darker than the surrounding. This needs further investigation.

A final OCT-phantom consisting of four different parameter test patterns (PSF, sensitivity, lateral resolution and distortion) was fabricated using *Low-Rep 1026 nm* fs laser system. A total of eight layers were inscribed with $75 \mu\text{m}$ spacing between them starting at $100 \mu\text{m}$ below surface. The PSF, lateral resolution, and distortion were inscribed with 50% of the maximum energy, that produced $4.8 \mu\text{m}$ (lateral width) structures. The Sensitivity Pattern was inscribed with varying energy. The PSF and Sensitivity patterns both have a pitch (same layer) of $150 \mu\text{m}$, while the Distortion Pattern comprises of a $100 \mu\text{m} \times 100 \mu\text{m}$ grids. The line spacing in the lateral resolution pattern varies linearly from $1 \mu\text{m}$ to $50 \mu\text{m}$. Inscribed structures were characterized using an optical microscope. Furthermore, the performance of the OCT-phantom was verified by several OCT systems, including MD-OCT, Santec-OCT, home built RT-OCT and several current Thorlab OCT systems.)

Quantitative methods have been developed to measure the performance of an OCT system. Measurement possibilities include spatially varying PSF and sensitivity. Quantitative methods for measuring and correcting distortion are under development. A method was developed to measure the spatially varying PSF of

the OCT system using a PSF pattern (OCT-phantom). This method produces a map of PSF widths as function of two or three coordinates. Although the description given here were for 2D (i.e. for B-scan), the method can be extended for 3D as well.

Image enhancement by deconvolution was demonstrated using the measured PSF values. The GPU based LR and WF deconvolution were applied to the OCT data, in linear scale. The deconvolution was carried out in three different ways: using a single average PSF for entire image plane; splitting the image and then deconvolving each tile using a single average PSF, and deconvolving each tile using its corresponding PSF values. The results shows WF routing is highly sensitive to the supplied noise to signal ratio (NSR), in this calculation we used a value that gave optimal results. Experimentally measuring the NSR might improve the results further. Quantitative improvement after applying deconvolution was evaluated by measuring LC, CNR, average SNR (SNR_{av}) and sharpness (K). The quantitative analysis confirms the improvements of image quality after deconvolution. For all three methods, LR achieves better results than WF.

A GPU based processing method was developed to accelerate the standard OCT data processing, this includes: resampling the data from linear in wavelength to linear in wave-number, dividing by reference spectrum, removing background, applying windowing functions calculating FFTs, modulus and converting to log scale (dB). Interpolation of the data using a reference table was carried out by resampling using linear or third order Lagrangian integrations routines. This implementation exploits several resources available within GPU such as different memory types, concurrent copy and execution to achieve maximum performance. A processing rate of 1.8 million A-scans per second was achieved.

The FWHM axial and lateral width of the structures was estimated by fitting a one dimension (1D) or 2D Gaussian function to the measured data. Gaussian fitting routines were implemented CPU-GPU hybrid to accelerate the processing speed. In addition, a description of estimating the refractive index from phase data measured by QPM is also given.

Two deconvolution routines, LR and WF were implemented in GPU. These implementations were specifically designed for tiled deconvolution. For this, the B-scan is split into small tiles, where each tile was deconvolved using its corresponding PSF.

An SLED based SB-OCT system was developed. This uses an SLED as the light source, a line scan detector as the detector, and a pair of galvo mirrors for scanning. Software was developed to acquire, process and render the data and control the system. The GPU based OCT data processing was adopted to process the data in real time. The current version of this system captures and processes the data at around 92 kHz , which is limited by the acquisition speed of the camera.

This thesis describes the production of standardization phantoms for validating the performance of OCT systems and the acceleration of the processing speeds using GPUs. Producing phantoms and developing quantitative methods has been a difficult task. Several potential solutions have been developed with varying degrees of success. Several other methods (phantom designs and quantitative characterization) are in early stages of the development process. This work will serve as a foundation for extending the development of phantoms and standardization methods.

11.2 Future work

The work presented in this thesis demonstrates the potential of various technologies and techniques for producing OCT calibration and characterization phantoms, and to improve the data processing rate. However, this work can be extended further. The following section gives a short discussion of some of the future work directions that could be approached to improve the quality of the phantoms, characterization methods and data processing.

11.2.1 Parameter space study

There is a delicate balance between the parameters used for producing small structures and observing them in OCT. Further investigation of the parameter space study could reveal the possibility of reducing the feature size whilst maintaining the visibility under OCT.

The current sample is inscribed with 50% energy and exhibits a lateral size of $4.8 \mu\text{m}$. Our investigation suggested that it may be possible to produce smaller structures if the right balance can be achieved. Future phantoms should be produced with lower energy, which produce smaller structures than the resolution of the OCT ($< 2 \mu\text{m}$) hence allows direct measurement of the PSF/MTF.

At 1 kHz repetition rate (using *Low-Rep 800 nm fs* laser system) I was able to produce pearl chain like structures in fused silica, borosilicate and PMMA. These structures were highly reproducible. Refractive index and cross sectional morphology of these structures have not been studied so far and they warrant further investigation.

All the samples produced by the *Low-Rep 1026 nm* system were inscribed at the repetition rate of 100 kHz . Future studies would ideally investigate the properties of the structures produced when varying the repetition rate of this system. At lower rates, we might be able to produce smooth pearl chain like structures as seen with *Low-Rep 800 nm* system.

In this study, we inscribed structures several hundred microns in depth. The results showed, that for constant energy, the size of the structure decreased when increasing depth. In future, we could try to minimize the variation of the size with depth by increasing the energy with depth.

11.2.2 Phantom design and fabrication

Current phantom designs allows to qualitatively and quantitatively determine many of the permanence metrics of the OCT. Further development of different designs should be undertaken to investigate other parameters such as birefringence. Fur-

ther work is also required on composite phantoms (i.e. laser inscription inside resin phantoms with varying density).

11.2.3 Quantitative methods for characterization

A variety of quantitative methods could be developed to quantify the performances of the system using the current and the future phantoms.

For example, currently we are developing algorithms that automatically measure the distortion of the grid phantom (OCT-phantom) and also correction routines that compensate the distortion of the OCT image (2D/3D) based on the measured distortion values. A routine for estimating the distortion of an OCT using our current phantom (distortion pattern) is currently under development. Methods to automatically locate the positions of the control points (intersections) in image coordinate are being developed. These could also be used to calculate the deviation of these points from undistorted reference points.

It is worth investigating different resampling methods (or transfer function) for the correction of distortion in OCT images. We are currently investigating a neural network based approach for estimating scan linearity using PSF/Distortion patterns. This neural network based approach can be extended to deconvolution and distortion correction as well.

11.2.4 Software/Algorithm developments

Further advanced GPU based processing methods can be developed. A prototype of distortion measurement using our OCT-phantom and correction are under development in Matlab. In future, these should be migrated to GPU for real time distortion correction. In addition to this, other numerical routines can be developed and implemented to improve the quality of the OCT image such as noise reduction and image segmentation.

In the current implementation of RT-OCT system, the processed data is copied back to CPU for rendering using Microsoft visual studio libraries. The next step

CHAPTER 11. CONCLUSION

with this system is to develop an extensive GPU based rendering technique that is capable of rendering the data in 3D from the GPU in real time.

As a long-term project, it would be interesting to port the NVIDIA GPU based software to GPUs from other vendors (e.g. ATI Radeon). We have benchmarked the performance of different NVIDIA GPUs, and an extension of this it would be to benchmark test GPUs and FPGAs from other vendors.

References

- [1] P. H. Tomlins, G. N. Smith, P. D. Woolliams, J. Rasakanthan, and K. Sugden, "Femtosecond laser micro-inscription of optical coherence tomography resolution test artifacts," Biomed. Opt. Express, vol. 2, pp. 1319–1327, May 2011.
- [2] K. Takada, I. Yokohama, K. Chida, and J. Noda, "New measurement system for fault location in optical waveguide devices based on an interferometric technique," Appl. Opt., vol. 26, pp. 1603–1606, May 1987.
- [3] R. C. Youngquist, S. Carr, and D. E. N. Davies, "Optical coherence-domain reflectometry: a new optical evaluation technique," Opt. Lett., vol. 12, pp. 158–160, Mar 1987.
- [4] A. F. Fercher, K. Mengedoht, and W. Werner, "Eye-length measurement by interferometry with partially coherent light," Opt. Lett., vol. 13, pp. 186–188, Mar 1988.
- [5] C. K. Hitzenberger, "Measurement of corneal thickness by low-coherence interferometry," Appl. Opt., vol. 31, pp. 6637–6642, Nov 1992.
- [6] J. A. Izatt, M. R. Hee, G. M. Owen, E. A. Swanson, and J. G. Fujimoto, "Optical coherence microscopy in scattering media," Optics Letters, vol. 19, pp. 590–592, April 1994.

REFERENCES

- [7] D. Huang, E. Swanson, C. Lin, and J. Schuman, "Optical coherence tomography," Science, vol. 254, pp. 1178–1181, November 1991.
- [8] A. Fercher, C. Hitzenberger, G. Kamp, and S. El-Zaiat, "Measurement of intraocular distances by backscattering spectral interferometry," Optics Communications, vol. 117, no. 1–2, pp. 43 – 48, 1995.
- [9] B. Potsaid, I. Gorczynska, V. J. Srinivasan, Y. Chen, J. Jiang, A. Cable, and J. G. Fujimoto, "Ultrahigh speed spectral / fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second," Opt. Express, vol. 16, pp. 15149–15169, Sep 2008.
- [10] B. Potsaid, B. Baumann, D. Huang, S. Barry, A. E. Cable, J. S. Schuman, J. S. Duker, and J. G. Fujimoto, "Ultrahigh speed 1050nm swept source/-fourier domain OCT retinal and anterior segment imaging at 100,000 to 400,000 axial scans per second," Opt. Express, vol. 18, pp. 20029–20048, Sep 2010.
- [11] R. Huber, D. C. Adler, and J. G. Fujimoto, "Buffered fourier domain mode locking: Unidirectional swept laser sources for optical coherence tomography imaging at 370,000 lines/s," Opt Lett, vol. 31, pp. 2975–2977, Oct 2006.
- [12] V. J. Srinivasan, D. C. Adler, Y. Chen, I. Gorczynska, R. Huber, J. S. Duker, J. S. Schuman, and J. G. Fujimoto Investigative Ophthalmology and Visual Science.
- [13] I. Grulkowski, M. Gora, M. Szkulmowski, I. Gorczynska, D. Szlag, S. Marcos, A. Kowalczyk, and M. Wojtkowski, "Anterior segment imaging with spectral OCT system using a high-speed CMOS camera," Opt. Express, vol. 17, pp. 4842–4858, Mar 2009.
- [14] G. C. B. Lee, J. Rasakanthan, P. D. Woolliams, and K. Sugden, "Fabrication

REFERENCES

- of high quality optical coherence tomography (OCT) calibration artefacts using femtosecond inscription,” vol. 8427, p. 84271K, Proc. SPIE, 2012.
- [15] J. Rasakanthan, M. Baregheh, M. Dubov, V. Mezentsev, G. Smith, K. Sugden, P. D. Woolliams, and P. H. Tomlins, “Quantitative optical coherence tomography for characterization of microscopic structures with varying refractive index,” in CLEO/Europe and EQEC 2011 Conference Digest, p. CH3 3, Optical Society of America, 2011.
- [16] P. Tomlins and R. Wang, “Theory, developments and applications of optical coherence tomography,” Journal of Physics D-Applied Physics, vol. 38, no. 15, pp. 2519–2535, 2005.
- [17] X. Clivaz, F. Marquis-Weible, R. P. Salathé, R. P. Novák, and H. H. Gilgen, “High-resolution reflectometry in biological tissues,” Opt. Lett., vol. 17, pp. 4–6, Jan 1992.
- [18] J. M. Schmitt, A. Knüttel, and R. F. Bonner, “Measurement of optical properties of biological tissues by low-coherence reflectometry,” Appl. Opt., vol. 32, pp. 6032–6042, Oct 1993.
- [19] A. Gh. Podoleanu, “Optical coherence tomography,” Journal of Microscopy, vol. 247, pp. 209–219, Sep 2012.
- [20] M. Wojtkowski, R. Leitgeb, A. Kowalczyk, T. Bajraszewski, and A. F. Fercher, “In vivo human retinal imaging by fourier domain optical coherence tomography,” J. Biomed. Opt., vol. 7, p. 457, July 2002.
- [21] N. Nassif, B. Cense, B. H. Park, S. H. Yun, T. C. Chen, B. E. Bouma, G. J. Tearney, and J. F. de Boer, “In vivo human retinal imaging by ultrahigh-speed spectral domain optical coherence tomography,” Opt. Lett., vol. 29, pp. 480–482, Mar 2004.

REFERENCES

- [22] B. Cense, N. Nassif, T. Chen, M. Pierce, S.-H. Yun, B. Park, B. Bouma, G. Tearney, and J. de Boer, "Ultra-high-resolution high-speed retinal imaging using spectral-domain optical coherence tomography," Opt. Express, vol. 12, pp. 2435–2447, May 2004.
- [23] M. Wojtkowski, V. Srinivasan, T. Ko, J. Fujimoto, A. Kowalczyk, and J. Duker, "Ultra-high-resolution, high-speed, fourier domain optical coherence tomography and methods for dispersion compensation," Opt. Express, vol. 12, pp. 2404–2422, May 2004.
- [24] S. R. Chinn, E. A. Swanson, and J. G. Fujimoto, "Optical coherence tomography using a frequency-tunable optical source," Opt. Lett., vol. 22, pp. 340–342, Mar 1997.
- [25] S. Yun, G. Tearney, J. de Boer, N. Iftimia, and B. Bouma, "High-speed optical frequency-domain imaging," Opt. Express, vol. 11, pp. 2953–2963, Nov 2003.
- [26] M. Choma, M. Sarunic, C. Yang, and J. Izatt, "Sensitivity advantage of swept source and fourier domain optical coherence tomography," Opt. Express, vol. 11, pp. 2183–2189, Sep 2003.
- [27] R. Leitgeb, C. Hitzenberger, and A. Fercher, "Performance of fourier domain vs. time domain optical coherence tomography," Opt. Express, vol. 11, pp. 889–894, Apr 2003.
- [28] Optical Coherence Tomography: Technology and Applications (Biological and Medical Physics, Biomedical Engineering). Springer, 2008.
- [29] J. Izatt, M. Kulkarni, H.-W. Wang, K. Kobayashi, and J. Sivak, M.V., "Optical coherence tomography and microscopy in gastrointestinal tissues," Selected Topics in Quantum Electronics, IEEE Journal of, vol. 2, pp. 1017–1028, Dec 1996.

REFERENCES

- [30] E. Brigham, The Fast Fourier Transform and Its Applications. Prentice-Hall Signal Processing Series, Prentice Hall, 1 ed., March 1988.
- [31] P. Tomlins, R. Ferguson, C. Hart, and P.D. Woolliams, "Point-spread function phantoms for optical coherence tomography," NPL Report (OP2), vol. ISSN: 1754-2944, 2009.
- [32] H. Lubatschowski, A. Heisterkamp, F. Will, J. Serbin, T. Bauer, C. Fallnich, H. Welling, W. Mueller, B. Schwab, A. I. Singh, and W. Ertmer, "Ultrafast laser pulses for medical applications," Commercial and Biomedical Applications of Ultrafast and Free-Electron Lasers, vol. 4633, pp. 38–49, Jan 2002.
- [33] K. M. Davis, K. Miura, N. Sugimoto, and K. Hirao, "Writing waveguides in glass with a femtosecond laser," Opt. Lett., vol. 21, pp. 1729–1731, Nov 1996.
- [34] A. M. Streltsov and N. F. Borrelli, "Fabrication and analysis of a directional coupler written in glass by nanojoule femtosecond laser pulses," Opt. Lett., vol. 26, pp. 42–43, Jan 2001.
- [35] Y. Cheng, K. Sugioka, K. Midorikawa, M. Masuda, K. Toyoda, M. Kawachi, and K. Shihoyama, "Three-dimensional micro-optical components embedded in photosensitive glass by a femtosecond laser," Opt. Lett., vol. 28, pp. 1144–1146, Jul 2003.
- [36] R. S. Taylor, C. Hnatovsky, E. Simova, D. M. Rayner, V. R. Bhardwaj, and P. B. Corkum, "Femtosecond laser fabrication of nanostructures in silica glass," Opt. Lett., vol. 28, pp. 1043–1045, Jun 2003.
- [37] V. Mezentsev, M. Dubov, A. Martinez, Y. Lai, T. P. Allsop, I. Khrushchev, D. J. Webb, F. Floreani, and I. Bennion, "Micro-fabrication of advanced photonic devices by means of direct point-by-point femtosecond inscription in silica," Laser-based Micropackaging, vol. 6107, no. 1, p. 61070C, 2006.

REFERENCES

- [38] D. Homoelle, S. Wielandy, A. L. Gaeta, N. F. Borrelli, and C. Smith, "Infrared photosensitivity in silica glasses exposed to femtosecond laser pulses," Opt. Lett., vol. 24, pp. 1311–1313, Sep 1999.
- [39] C. B. Schaffer, Interaction of Femtosecond Laser Pulses with Transparent Materials. PhD thesis, The Department of Physics, Harvard University, 2001.
- [40] K. Miura, J. Qiu, H. Inouye, T. Mitsuyu, and K. Hirao, "Photowritten optical waveguides in various glasses with ultrashort pulse laser," Appl. Phys. Lett., vol. 71, no. 23, pp. 3329–3331, 1997.
- [41] K. Miura, J. Qiu, T. Mitsuyu, and K. Hirao, "Three-dimensional microscopic modifications in glasses by a femtosecond laser," vol. 3618, pp. 141–151, Proc. SPIE, Jan 1999.
- [42] A. M. Streltsov and N. F. Borrelli, "Study of femtosecond-laser-written waveguides in glasses," J. Opt. Soc. Am. B, vol. 19, pp. 2496–2504, Oct 2002.
- [43] F. Korte, S. Adams, A. Egbert, C. Fallnich, A. Ostendorf, S. Nolte, M. Will, J.-P. Ruske, B. Chichkov, and A. Tünnemann, "Sub-diffraction limited structuring of solid targets with femtosecond laser pulses," Opt. Express, vol. 7, pp. 41–49, Jul 2000.
- [44] C. Valdivia, X. M. Wei, D. Coric, and P. R. Herman, "F2 laser-induced visible- and infrared-confining buried waveguides in fused silica," Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science Conference, p. CWI4, Jun 2003.
- [45] M. Will, S. Nolte, B. N. Chichkov, and A. Tünnemann, "Optical properties of waveguides fabricated in fused silica by femtosecond laser pulses," Appl. Opt., vol. 41, pp. 4360–4364, Jul 2002.

REFERENCES

- [46] K. Minoshima, A. M. Kowalewicz, I. Hartl, E. P. Ippen, and J. G. Fujimoto, "Photonic device fabrication in glass by use of nonlinear materials processing with a femtosecond laser oscillator," Opt. Lett., vol. 26, pp. 1516–1518, Oct 2001.
- [47] C. B. Schaffer, J. F. Garcia, and E. Mazur, "Bulk heating of transparent materials using a high repetition-rate femtosecond laser," Appl. Phys. A, vol. 76, pp. 351–354, Mar 2003.
- [48] Y. Sikorski, A. Said, P. Bado, R. Maynard, C. Florea, and K. Winick, "Optical waveguide amplifier in Nd-doped glass written with near-IR femtosecond laser pulses," Electron. Lett., vol. 36, no. 3, 2000.
- [49] S. Taccheo, G. D. Valle, R. Osellame, G. Cerullo, N. Chiodo, P. Laporta, O. Svelto, A. Killi, U. Morgner, M. Lederer, and D. Kopf, "Er:Yb-doped waveguide laser fabricated by femtosecond laser pulses," Opt. Lett., vol. 29, pp. 2626–2628, Nov 2004.
- [50] E. N. Glezer, M. Milosavljevic, L. Huang, R. J. Finlay, T.-H. Her, J. P. Callan, and E. Mazur, "Three-dimensional optical storage inside transparent materials," Opt. Lett., vol. 21, pp. 2023–2025, Dec 1996.
- [51] Y. Kondo, K. Nouchi, T. Mitsuyu, M. Watanabe, P. G. Kazansky, and K. Hiro, "Fabrication of long-period fiber gratings by focused irradiation of infrared femtosecond laser pulses," Opt. Lett., vol. 24, pp. 646–648, May 1999.
- [52] L. Sudrie, M. Franco, B. Prade, and A. Mysyrowicz Optics Communications.
- [53] X. Liu, D. Du, and G. Mourou, "Laser ablation and micromachining with ultrashort laser pulses," Quantum Electronics, IEEE Journal of, vol. 33, pp. 1706–1716, Oct 1997.

REFERENCES

- [54] Y. Bellouard, A. Said, M. Dugan, and P. Bado, "Fabrication of high-aspect ratio, micro-fluidic channels and tunnels using femtosecond laser pulses and chemical etching," Opt. Express, vol. 12, May 2004.
- [55] A. Saliminia, S. L. Chin, and R. Vallée, "Ultra-broad and coherent white light generation in silica glass by focused femtosecond pulses at 1.5 μm ," Opt. Express, vol. 13, pp. 5731–5738, Jul 2005.
- [56] A. Brodeur and S. L. Chin, "Ultrafast white-light continuum generation and self-focusing in transparent condensed media," J. Opt. Soc. Am. B, vol. 16, pp. 637–650, Apr 1999.
- [57] S. Tzortzakis, L. Sudrie, M. Franco, B. Prade, A. Mysyrowicz, A. Couairon, and L. Bergé, "Self-guided propagation of ultrashort IR laser pulses in fused silica," Phys. Rev. Lett., vol. 87, p. 213902, Nov 2001.
- [58] G. Yang and Y. R. Shen, "Spectral broadening of ultrashort pulses in a nonlinear medium," Opt. Lett., vol. 9, pp. 510–512, Nov 1984.
- [59] V. Mezentsev, J. S. Petrovic, M. Dubov, I. Bennion, J. Dreher, H. Schmitz, and R. Grauer, "Femtosecond laser microfabrication of subwavelength structures in photonics," Laser-based Micro- and Nanopackaging and Assembly, vol. 6459, no. 1, p. 64590B, 2007.
- [60] Q. Xing, K. M. Yoo, and R. R. Alfano, "Conical emission by four-photon parametric generation by using femtosecond laser pulses," Appl. Opt., vol. 32, pp. 2087–2089, Apr 1993.
- [61] W. J. Reichman, Understanding how femtosecond laser waveguide fabrication in glasses works. PhD thesis, UNIVERSITY OF CALIFORNIA, 2006.
- [62] W. Yang, Femtosecond laser writing in transparent materials. PhD thesis, Optoelectronics Research Centre, Faculty of Engineering, Science and Mathematics, University of Southampton, 2008.

REFERENCES

- [63] S. C. Jones, P. Braunlich, R. T. Casper, X.-A. Shen, and P. Kelly, "Recent progress on laser-induced modifications and intrinsic bulk damage of wide-gap optical materials," Optical Engineering, vol. 28, pp. 1039–1068, Oct. 1989.
- [64] B. C. Stuart, M. D. Feit, S. Herman, A. M. Rubenchik, B. W. Shore, and M. D. Perry, "Nanosecond-to-femtosecond laser-induced breakdown in dielectrics," Phys. Rev. B, vol. 53, pp. 1749–1761, Jan 1996.
- [65] D. Du, X. Liu, G. Korn, J. Squier, and G. Mourou, "Laser-induced breakdown by impact ionization in SiO_2 with pulse widths from 7 ns to 150 fs," Appl. Phys. Lett., vol. 64, no. 23, pp. 3071–3073, 1994.
- [66] K. Sharma, Optics: Principles And Applications. Academic Press, 2006.
- [67] R. Suo, Advanced fibre gratings fabricated in standard and infrared glass fibres by ultraviolet and near-infrared-femtosecond lasers. PhD thesis, Aston University, 2009.
- [68] P. Yu, Fundamentals of Semiconductors. Heidelberg, Germany: Springer-Verlag, 1996.
- [69] E. N. Glezer and E. Mazur, "Ultrafast-laser driven micro-explosions in transparent materials," Appl. Phys. Lett., vol. 71, pp. 882–884, June 1997.
- [70] B. C. Stuart, M. D. Feit, A. M. Rubenchik, B. W. Shore, and M. D. Perry, "Laser-induced damage in dielectrics with nanosecond to subpicosecond pulses," Phys. Rev. Lett., vol. 74, pp. 2248–2251, Mar 1995.
- [71] R. Osellame, S. Taccheo, M. Marangoni, R. Ramponi, P. Laporta, D. Polli, S. D. Silvestri, and G. Cerullo, "Femtosecond writing of active optical waveguides with astigmatically shaped beams," J. Opt. Soc. Am. B, vol. 20, pp. 1559–1567, Jul 2003.

REFERENCES

- [72] B. C. Stuart, M. D. Feit, S. Herman, A. M. Rubenchik, B. W. Shore, and M. D. Perry, "Optical ablation by high-power short-pulse lasers," J. Opt. Soc. Am. B, vol. 13, pp. 459–468, Feb 1996.
- [73] V. Mezentsev, S. Turitsyn, M. Fedoruk, M. Dubov, A. Rubenchik, and E. Podivilov, "Modelling of femtosecond inscription in fused silica," in Transparent Optical Networks, 2006 International Conference on, vol. 4, pp. 146 –149, June 2006.
- [74] X. Wang, J. Qiu, J. Song, J. Xu, Y. Liao, H. Sun, Y. Cheng, and Z. Xu, "Simultaneous three-photon absorption induced ultraviolet upconversion in $pr^{3+} : y_2sio_5$ crystal by femtosecond laser irradiation," Optics Communications, vol. 281, pp. 299–302, Jan 2008.
- [75] E. Yablonovitch and N. Bloembergen, "Avalanche ionization and the limiting diameter of filaments induced by light pulses in transparent media," Phys. Rev. Lett., vol. 29, pp. 907–910, Oct 1972.
- [76] D. Du, X. Liu, and G. Mourou, "Reduction of multi-photon ionization in dielectrics due to collisions," Applied Physics B: Lasers and Optics, vol. 63, pp. 617–621, 1996. 10.1007/BF01831002.
- [77] T. Q. Jia, H. X. Chen, M. Huang, F. L. Zhao, X. X. Li, S. Z. Xu, H. Y. Sun, D. H. Feng, C. B. Li, X. F. Wang, R. X. Li, Z. Z. Xu, X. K. He, and H. Kuroda, "Ultraviolet-infrared femtosecond laser-induced damage in fused silica and caf_2 crystals," Phys. Rev. B, vol. 73, p. 054105, Feb 2006.
- [78] J. H. Marburger, "Self-focusing: theory," Prog. Quantum Electron., vol. 4, no. 35-110, 1975.
- [79] R. W. Boyd p. 576, Jan 2003.
- [80] A. L. Gaeta, "Catastrophic collapse of ultrashort pulses," Phys. Rev. Lett., vol. 84, pp. 3582–3585, Apr 2000.

REFERENCES

- [81] M. Ams, G. Marshall, D. Spence, and M. Withford, "Slit beam shaping method for femtosecond laser direct-write fabrication of symmetric waveguides in bulk glasses," Opt. Express, vol. 13, pp. 5676–5681, Jul 2005.
- [82] A. Saliminia, N. Nguyen, M. Nadeau, and S. Petit, "Writing optical waveguides in fused silica using 1 kHz femtosecond infrared pulses," Journal of Applied Physics, Jan 2003.
- [83] A. Joglekar, H. Liu, G. Spooner, E. Meyhöfer, G. Mourou, and A. Hunt, "A study of the deterministic character of optical damage by femtosecond laser pulses and applications to nanomachining," Applied Physics B: Lasers and Optics, vol. 77, pp. 25–30, Jul 2003.
- [84] T. Klein, W. Wieser, L. Reznicek, A. Neubauer, A. Kampik, and R. Huber, "Multi-MHz retinal OCT," Biomed. Opt. Express, vol. 4, pp. 1890–1908, Oct 2013.
- [85] R. Huber, K. Taira, M. Wojtkowski, T. H. Ko, J. G. Fujimoto, and K. Hsu, "High-speed frequency swept light source for fourier domain oct at 20 khz a-scan rate," Proc. SPIE, vol. 5690, pp. 96–100, 2005.
- [86] M. Gora, K. Karnowski, M. Szkulmowski, B. J. Kaluzny, R. Huber, A. Kowalczyk, and M. Wojtkowski, "Ultra high-speed swept source OCT imaging of the anterior segment of human eye at 200 khz with adjustable imaging range," Opt. Express, vol. 17, pp. 14880–14894, Aug 2009.
- [87] M. Gargesha, M. W. Jenkins, D. L. Wilson, and A. M. Rollins, "High temporal resolution OCT using image-based retrospective gating," Opt. Express, vol. 17, pp. 10786–10799, Jun 2009.
- [88] M. Gargesha, M. W. Jenkins, A. M. Rollins, and D. L. Wilson, "Denoising and 4D visualization of OCT images," Opt. Express, vol. 16, pp. 12313–12333, Aug 2008.

REFERENCES

- [89] M. W. Jenkins, F. Rothenberg, D. Roy, V. P. Nikolski, Z. Hu, M. Watanabe, D. L. Wilson, I. R. Efimov, and A. M. Rollins, "4D embryonic cardiography using gated optical coherence tomography," Opt. Express, vol. 14, pp. 736–748, Jan 2006.
- [90] K. Zhang and J. U. Kang, "Real-time 4D signal processing and visualization using graphics processing unit on a regular nonlinear-k fourier-domain OCT system," Opt. Express, vol. 18, pp. 11772–11784, May 2010.
- [91] Y. Zhang, X. Li, L. Wei, K. Wang, Z. Ding, and G. Shi, "Time-domain interpolation for fourier-domain optical coherence tomography," Opt Lett, vol. 34, pp. 1849–1851, Jan 2009.
- [92] A. Desjardins, B. Vakoc, M. Suter, S.-H. Yun, G. Tearney, and B. Bouma, "Real-time fpga processing for high-speed optical frequency domain imaging," Medical Imaging, IEEE Transactions on, vol. 28, pp. 1468 –1472, sept. 2009.
- [93] J. Li, M. Sarunic, and L. Shannon, "Scalable, high performance fourier domain optical coherence tomography: Why FPGAs and not GPGPUs," in Field-Programmable Custom Computing Machines (FCCM), 2011 IEEE 19th Annual International Symposium on, pp. 49 –56, May 2011.
- [94] A. W. Schaefer, J. J. Reynolds, D. L. Marks, and S. A. Boppart, "Real-time digital signal processing-based optical coherence tomography and doppler optical coherence tomography," IEEE Trans Biomed Eng, vol. 51, pp. 186–190, Jan 2004.
- [95] S. Yan, D. Piao, Y. Chen, and Q. Zhu, "Digital signal processor-based real-time optical doppler tomography system," J. Biomed Opt., vol. 9, pp. 454–463, May-Jun 2004.
- [96] J. Su, J. Zhang, L. Yu, H. G. Colt, M. Brenner, and Z. Chen, "Real-time swept source optical coherence tomography imaging of the human airway

REFERENCES

- using a microelectromechanical system endoscope and digital signal processor,” J. Biomed Opt., vol. 13, no. 3, p. 030506, 2008.
- [97] G. Liu, J. Zhang, L. Yu, T. Xie, and Z. Chen, “Real-time polarization-sensitive optical coherence tomography data processing with parallel computing,” Appl. Opt., vol. 48, pp. 6365–6370, Nov 2009.
- [98] Z. Hu and A. M. Rollins, “Fourier domain optical coherence tomography with a linear-in-wavenumber spectrometer,” Opt. Lett., vol. 32, pp. 3525–3527, Dec 2007.
- [99] Y. Watanabe and T. Itagaki, “Real-time display on fourier domain optical coherence tomography system using a graphics processing unit,” J. Biomed Opt., Jan 2009.
- [100] S. V. der Jeught, A. Bradu, and A. G. Podoleanu, “Real-time resampling in fourier domain optical coherence tomography using a graphics processing unit,” J. Biomed Opt., vol. 15, no. 3, p. 030511, 2010.
- [101] J. Rasakanthan, K. Sugden, and P. H. Tomlins, “Processing and rendering of fourier domain optical coherence tomography images at a line rate over 524 kHz using a graphics processing unit,” J. Biomed Opt., vol. 16, no. 2, p. 020505, 2011.
- [102] K. Zhang and J. U. Kang, “Graphics processing unit accelerated non-uniform fast fourier transform for ultrahigh-speed, real-time fourier-domain OCT,” Opt. Express, vol. 18, pp. 23472–23487, Oct 2010.
- [103] X. Li, G. Shi, and Y. Zhang, “High-speed optical coherence tomography signal processing on GPU,” Journal of Physics: Conference Series, vol. 277, no. 1, p. 012019, 2011.
- [104] J. Probst, P. Koch, and G. Huttmann, “Real-time 3D rendering of optical coherence tomography volumetric data,” Optical Coherence Tomography and Coherence Techniques IV, vol. 7372, no. 1, p. 73720Q, 2009.

REFERENCES

- [105] B. R. Biedermann, W. Wieser, C. M. Eigenwillig, G. Palte, D. C. Adler, V. J. Srinivasan, J. G. Fujimoto, and R. Huber, “Real time en face fourier-domain optical coherence tomography with direct hardware frequency demodulation,” Opt Lett, vol. 33, pp. 2556–2558, Jan 2008.
- [106] NVIDIA, “Nvidia CUDA compute unified device architecture programming guide version 2.3.1,” tech. rep., 2009.
- [107] “Nvidia cuda zone (web),” 2012.
- [108] P. Maciol and K. Banas, “Testing tesla architecture for scientific computing: The performance of matrix-vector product,” in Computer Science and Information Technology, 2008. IMCSIT 2008. International Multiconference on, pp. 285 –291, Oct. 2008.
- [109] M. Hovorka, Training of neural networks using CUDA technology. PhD thesis, Czech Technical University in Prague, Faculty of Electrical Engineering, Department of Computer Science and Engineering, May 2010.
- [110] B. Dan. PhD thesis.
- [111] “CUDA C best practices guide,” tech. rep., 2012.
- [112] W. Drexler, “Ultrahigh-resolution optical coherence tomography,” J. Biomed Opt., vol. 9, pp. 47–74, Jan-Feb 2004.
- [113] C. Akcay, P. Parrein, and J. P. Rolland, “Estimation of longitudinal resolution in optical coherence imaging,” Appl. Opt., vol. 41, pp. 5256–5262, Sep 2002.
- [114] A. Agrawal, T. J. Pfefer, N. Gilani, and R. Drezek, “Three-dimensional characterization of optical coherence tomography point spread functions with a nanoparticle-embedded phantom,” Opt. Lett., vol. 35, pp. 2269–2271, Jul 2010.

REFERENCES

- [115] P. Tomlins, W. Palin, and A. Shortall, "Dynamic cure measurement of dental polymer composites using optical coherence tomography," Proceedings of SPIE, Jan 2008.
- [116] A. M. Davis, Development of fourier domain optical coherence tomography for applications in developmental biology. PhD thesis, Duke University, Jun 2008.
- [117] R. E. Ziemer and W. H. Tranter, Principles of Communications. Wiley, 1995.
- [118] A. Chennu, Optical Coherence Tomography A Literature Survey. PhD thesis, Heriot Watt University, 2008.
- [119] A. Fercher, W. Drexler, C. Hitzenberger, and T. Lasser Reports on progress in physics.
- [120] P. Woolliams, R. Ferguson, C. Hart, A. Grimwood, and P. Tomlins, "Spatially deconvolved optical coherence tomography," Applied Optics, vol. 49, pp. 2014–2021, April 2010.
- [121] R. C. Haskell, D. Liao, A. E. Pivonka, T. L. Bell, B. R. Haberle, B. M. Hoeling, and D. C. Petersen, "Role of beat noise in limiting the sensitivity of optical coherence tomography," J. Opt. Soc. Am. A, vol. 23, pp. 2747–2755, Nov 2006.
- [122] S. Ortiz, D. Siedlecki, I. Grulkowski, L. Remon, D. Pascual, M. Wojtkowski, and S. Marcos, "Optical distortion correction in optical coherence tomography for quantitative ocular anterior segment by three-dimensional imaging," Opt. Express, vol. 18, pp. 2782–2796, Jan 2010.
- [123] A. Podoleanu, I. Charalambous, L. Plesea, A. Dogariu, and R. Rosen, "Correction of distortions in optical coherence tomography imaging of the eye," Phys. Med. Biol., vol. 49, p. 1277, 2004.

REFERENCES

- [124] D. Wang, D. M. Doddrell, and G. Cowin, "A novel phantom and method for comprehensive 3-dimensional measurement and correction of geometric distortion in magnetic resonance imaging," Magn Reson Imaging, vol. 22, pp. 529–542, May 2004.
- [125] C. Fellner, W. Müller, J. Georgi, U. Taubenreuther, F. A. Fellner, and W. A. Kalender, "A high-resolution phantom for MRI," Magnetic Resonance Imaging, vol. 19, no. 6, pp. 899 – 904, 2001.
- [126] Z. Wang, C. E. Glazowski, and J. M. Zavislan, "Modulation transfer function measurement of scanning reflectance microscopes," J. Biomed Opt., vol. 12, no. 5, p. 051802, 2007.
- [127] P. D. Woolliams and P. H. Tomlins, "The modulation transfer function of an optical coherence tomography imaging system in turbid media," Phys. Med. Biol., vol. 56, no. 9, p. 2855, 2011.
- [128] P. D. Woolliams and P. H. Tomlins, "Estimating the resolution of a commercial optical coherence tomography system with limited spatial sampling," Measurement Science and Technology, vol. 22, no. 6, p. 065502, 2011.
- [129] Y. Liu, Y. Liang, G. Mu, and X. Zhu, "Deconvolution methods for image deblurring in optical coherence tomography," J. Opt. Soc. Am. A, vol. 26, pp. 72–77, Jan 2009.
- [130] T. van Leeuwen, D. Faber, and M. Aalders, "Measurement of the axial point spread function in scattering media using single-mode fiber-based optical coherence tomography," Selected Topics in Quantum Electronics, IEEE Journal of, vol. 9, pp. 227 – 233, Mar-Apr 2003.
- [131] M. Kulkarni, C. Thomas, and J. Izatt, "Image enhancement in optical coherence tomography using deconvolution," Electron. Lett., vol. 33, pp. 1365 –1367, Jul 1997.

REFERENCES

- [132] J. M. Schmitt and Z. Liang, "Deconvolution and enhancement of optical coherence tomograms," Coherence Domain Optical Methods in Biomedical Science and Clinical Applications, vol. 2981, no. 1, pp. 46–57, 1997.
- [133] K. Wang, Z. Ding, M. Chen, C. Wang, T. Wu, and J. Meng, "Deconvolution with fall-off compensated axial point spread function in spectral domain optical coherence tomography," Optics Communications, vol. 284, no. 12, pp. 3173–3180, 2011.
- [134] S. R. McNown and B. R. Hunt, "Approximate shift-invariance by warping shift-variant systems," in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (T. J. Schulz and D. L. Snyder, eds.), vol. 2302 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, pp. 156–167, Sep 1994.
- [135] J. G. Nagy and D. P. O’Leary, "Restoring images degraded by spatially variant blur," SIAM Journal on Scientific Computing, vol. 19, no. 4, pp. 1063–1082, 1998.
- [136] J. G. Nagy and D. P. O’Leary, "Restoring images degraded by spatially-variant blur," SIAM J. Sci. Comput, vol. 19, no. 4, pp. 1063–1082, 1997.
- [137] H. Trussell and S. Fogel, "Identification and restoration of spatially variant motion blurs in sequential images," Image Processing, IEEE Transactions on, vol. 1, pp. 123 –126, Jan 1992.
- [138] J. N. J. Bardsley, S. Jefferies and R. Plemmons, "Blind iterative restoration of images with spatially-varying blur," Opt. Express, vol. 14, pp. 1767–1782, 2006.
- [139] A. K. Jain, Fundamentals of Digital Image Processing,. Englewood Cliffs, NJ: Prentice-Hall, 1989.

REFERENCES

- [140] T. Allsop, M. Dubov, V. Mezentsev, and I. Bennion, "Inscription and characterization of waveguides written into borosilicate glass by a high-repetition-rate femtosecond laser at 800 nm," Appl. Opt., vol. 49, pp. 1938–1950, Apr 2010.
- [141] M. Dubov, V. Mezentsev, I. Bennion, and D. N. Nikogosyan, "UV femtosecond laser inscribes a 300 nm period nanostructure in a pure fused silica," Meas Sci Technol, vol. 18, pp. L15–L17, Jul 2007.
- [142] S. K. Turitsyn, V. K. Mezentsev, M. Dubov, A. M. Rubenchik, M. P. Fedoruk, and E. V. Podivilov, "Sub-critical regime of femtosecond inscription," Opt. Express, vol. 15, pp. 14750–14764, Oct 2007.
- [143] C. Florea and K. Winick, "Fabrication and characterization of photonic devices directly written in glass using femtosecond laser pulses," J. Lightwave Technol, vol. 21, pp. 246–253, 2003.
- [144] M. Yamaji, H. Kawashima, J. Suzuki, and S. Tanaka, "Three dimensional micromachining inside a transparent material by single pulse femtosecond laser through a hologram," Appl. Phys. Lett., vol. 93, p. 041116, Jul 2008.
- [145] J. Rasakanthan, G. C. B. Lee, P. D. Woolliams, and K. Sugden, "Parametric study of femtosecond inscription of microstructures for OCT artefact fabrication," vol. 8213, p. 82133N, Proc. SPIE, 2012.
- [146] P. Dekker, M. Ams, G. D. Marshall, D. J. Little, and M. J. Withford, "Annealing dynamics of waveguide bragg gratings: evidence of femtosecond laser induced colour centres," Opt. Express, vol. 18, pp. 3274–3283, Feb 2010.
- [147] S.-W. Lee, H.-W. Jeong, Y.-C. Ahn, W. Jung, Z. Chen, and B.-M. Kim, "Axial resolution and depth range of high-resolution spectral domain optical coherence tomography at 1.3 μm ," vol. 7168, p. 71682L, Proc. SPIE, 2009.

REFERENCES

- [148] K. Wang, Z. Ding, T. Wu, C. Wang, J. Meng, M. Chen, and L. Xu, "Development of a non-uniform discrete fourier transform based high speed spectral domain optical coherence tomography system," Opt. Express, vol. 17, pp. 12121–12131, Jul 2009.
- [149] A. Barty, K. A. Nugent, D. Paganin, and A. Roberts, "Quantitative optical phase microscopy," Opt. Lett., vol. 23, pp. 817–819, Jun 1998.
- [150] E. Ampem-Lassen, S. T. Huntington, N. M. Dragomir, K. A. Nugent, and A. Roberts, "Refractive index profiling of axially symmetric optical fibers: a new technique," Opt. Express, vol. 13, pp. 3277–3282, May 2005.
- [151] A. Barty, K. Nugent, A. Roberts, and D. Paganin, "Quantitative phase tomography," Optics Communications, vol. 175, no. 4–6, pp. 329 – 336, 2000.
- [152] M. Kalal and K. A. Nugent, "Abel inversion using fast fourier transforms," Appl. Opt., vol. 27, pp. 1956–1959, May 1988.
- [153] A. Roberts, E. Ampem-Lassen, A. Barty, K. A. Nugent, G. W. Baxter, N. M. Dragomir, and S. T. Huntington, "Refractive-index profiling of optical fibers with axial symmetry by use of quantitative phase microscopy," Opt. Lett., vol. 27, pp. 2061–2063, Dec 2002.
- [154] N. Dragomir, E. Ampem-Lassen, S. Huntington, G. Baxter, A. Roberts, and P. Farrell, "Refractive index profiling of optical fibers using differential interference contrast microscopy," Photonics Technology Letters, IEEE, vol. 17, pp. 2149 – 2151, Oct 2005.

Appendix A

Gaussian fit

This section describes the procedure used for estimating 2D Gaussian fit parameters. The algorithm used was based on Peter H Tomlins' MATLAB code. In order to increase the performance, fitting procedure is partially migrated to GPU.

The cross sectional profile of the femtosecond laser inscribed structures, at low energy, measured by an OCT can be modelled using a two dimensional Gaussian equation represented as:

$$G(x)_{2D} = Ae^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)} + g_0, \quad (\text{A.0.1})$$

where A is the amplitude, x_0 and y_0 are the x and y coordinates of the peak respectively, σ_x and σ_y are the FWHM axial and lateral width of the peak, and g_0 is the peak off-set from the image centre.

The two dimensional Gaussian profile given in the equation (A.0.1) estimates the axial w_x and lateral w_y FWHM widths, maximum peak intensity A and the position of the peak (x_0, y_0) on an ideal point from OCT B-scan. A small region of $x \times y$ pixels were chosen around each point from a B-scan data where the selected region must be large enough to contain the structure and small enough to avoid selecting neighbouring points. Two-dimensional Gaussian function was fitted using non-linear least square fitting method. Algorithms describe the fitting

APPENDIX A. GAUSSIAN FIT

procedure in details.

Measured values $g(x)$ in the sub region can be considered as a Gaussian function with fit parameters $G(x, y; A, x_0, y_0, \sigma_x, \sigma_y, g_0)$. Residue $d\beta_{i,j}$ is given by

$$d\beta_{i,j} = g_{i,j} - G(x_i, y_i, A, x_0, y_0, \sigma_x, \sigma_y, g_0), \quad (\text{A.0.2})$$

where subscripts i and j refer to the row and column indices of the two dimensional data.

The least square residue of an over determined equation¹ can be obtained by minimising the square of the residue in the 2-norm $\|r_i\|_2^2 = r^T r$. The first derivative of the residue is equal to zero $\frac{d\beta}{d\lambda} = 0$ at minimum, $d\mathbf{f} = Jd\lambda$, where $J \in \mathbb{R}^{m \times n}$ is the Jacobian matrix, which contains the partial derivatives of the function $G(x, y; \lambda)$. Partial derivatives of the function G with respect to all the fit parameters can be written as:

$$\begin{aligned} \frac{\partial G}{\partial A} &= \frac{G - g_0}{A} \\ \frac{\partial G}{\partial x_0} &= -2 \frac{x - x_0}{\sigma_x^2} (G - g_0) \\ \frac{\partial G}{\partial y_0} &= -2 \frac{y - y_0}{\sigma_y^2} (G - g_0) \\ \frac{\partial G}{\partial \sigma_x} &= -2 \frac{(x - x_0)^2}{\sigma_x^3} (G - g_0) \\ \frac{\partial G}{\partial \sigma_y} &= -2 \frac{(y - y_0)^2}{\sigma_y^3} (G - g_0) \\ \frac{\partial G}{\partial g_0} &= 1 \end{aligned} \quad (\text{A.0.3})$$

Singular value decomposition (SVD) is a decomposition method that can be used to decompose a matrix $J \in \mathbb{R}^{m \times n}$, where $m > n$ in to column orthogonal matrix $U \in \mathbb{R}^{m \times n}$, orthogonal matrix $U \in \mathbb{R}^{n \times n}$ and a diagonal matrix containing the singular values $S \in \mathbb{R}^{m \times n}$

¹system with more equations than the unknown

$$J = U\Sigma V^T. \quad (\text{A.0.4})$$

The shift in the fit parameter is given as $d\lambda = J_{pseudo}d\mathbf{f}$, where the Pseudo-inverse of a matrix J can be expressed as

$$J_{pseudo} = V\Sigma^*U^T, \quad (\text{A.0.5})$$

where Σ^* is the pseudo-inverse of the matrix Σ , in which all the non-zero elements are replaced by their reciprocal.

$$\Sigma_{i,j}^* = \begin{cases} \Sigma_{i,j}^{-1} & \Leftrightarrow (i = j) \wedge (\Sigma_{i,j} \neq 0) \\ 0 & otherwise \end{cases} \quad (\text{A.0.6})$$

The SVD routine available in CULA tools was used for decomposing the transpose of the Jacobian J^T into $U\Sigma V^T$. $J^k \in \mathbb{R}^{m \times n}$ where $k \in 0, 1, 2, 3, 4, 5$ represents the partial derivatives with respect to each of the fit parameters. All J^k are stored in a single continuous memory as follows:

$$\begin{aligned} J_{0:mn-1} &= J_{1:mn-1}^1 \\ J_{mn:2*mn-1} &= J_{1:mn-1}^2 \\ &\vdots \\ J_{kM:(k+1)M-1} &= J_{1:mn-1}^k \end{aligned} \quad (\text{A.0.7})$$

$$(\text{A.0.8})$$

The following section describes the implementation of this non-linear least square fit into the GPU-CPU hybrid architecture ². This fitting routine uses well optimised routine provided in the CUDA software development kit (SDK) for matrix

²Original routine was written in MATLAB by P.H.Tomlins

APPENDIX A. GAUSSIAN FIT

multiplication, transpose and addition. Computation of matrix multiplication of two matrices $A, B \in \mathbb{R}^{m \times n}$ was based on the CUDA SDK example, which uses Strassen's algorithm.

The iteration procedure starts with an initial set of guessed fit parameters λ_0 which will be improved over time. *Kernel 1* computes the Gaussian values $G(x, y; \lambda)$ using the fit parameters and then calculates the difference between the data and the model. *Kernel 2* computes the matrix J with respect to all fit parameters at every point in the matrix G , and takes the transpose of J

The SVD routine available in CULA was used for decomposing the matrix J^T into U, V^T and η (i.e. *Kernel 3*), where η is a vector of size $\min(m, n)$ that contains all the non-zero singular values of the diagonal matrix Σ . *Kernel 4* computes the reciprocal of the η which is used to create the pseudo inverse of the matrix Σ in *Kernel 5*. The transpose of the matrices U and V^T was used to calculate the $d\beta$ in *Kernel 5*.

In order to validate the correctness of the fitting routine a Gaussian peak with the $\hat{\lambda} = [1, 11, 11, 3, 3, 1]$ is produced with a white Gaussian noise of 20 dB. The fitting routine gives best estimate of the fit parameters for this peak as: $\hat{\lambda} = [1, 11, 11, 3, 3, 1]$. The noise Gaussian and the regenerated Gaussian from the estimated fit parameters are given in Figure A.1.

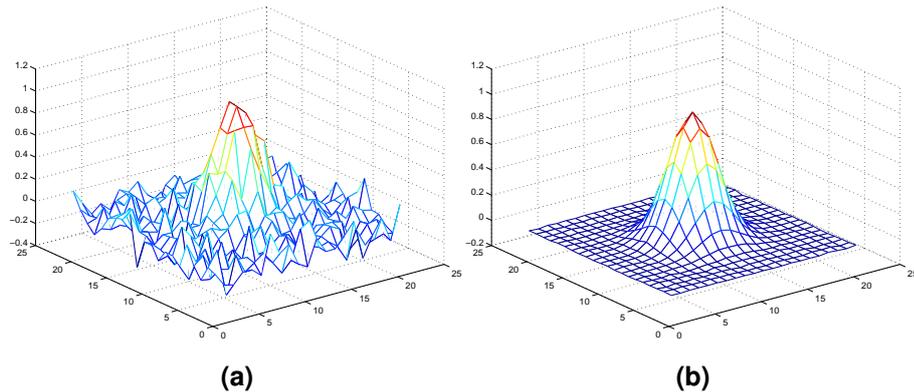


Figure A.1: Simulated input data (a) and the peak estimated by fitting a 2D Gaussian (b). Input data was simulated with the fit parameters $\{1, 11, 11, 3, 3, 1\}$ and with a 20 dB noise.

Appendix B

Reconstructing the refractive index

QPM is a non-interferometric technique that uses partially coherent light in a bright field microscope, that can determine the phase shift introduced in wave field in transparent materials [149].

Propagation of light through the modified regions along the z-axis of the microscope obeys the transport of intensity equation [150, 151]:

$$\frac{2\pi}{\lambda} \frac{\partial I(\hat{r})}{\partial z} = -\nabla \cdot \left(I(\hat{r}) \nabla \phi \left(\frac{\hat{r}}{M} \right) \right), \quad (\text{B.0.1})$$

where $r = (x, y)$ is a two dimensional vector in the transverse plane, I is the intensity distribution of the image and M is the magnification of the microscope.

APPENDIX B. RECONSTRUCTING THE REFRACTIVE INDEX

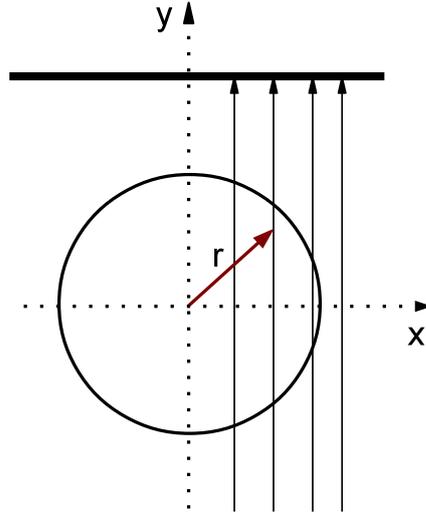


Figure B.1: Illustrating the propagation of light through a cylindrical structure of radius r .

A transmission optical microscope with a $20\times$ objective was used to obtain a set of images; one in focus (I_0) and others slightly defocused images either side of the focused image (I_{\pm}). The defocused images were obtained by displacing the sample in the positive and negative direction (i.e. $\pm z$) using the translating stage in the microscope. A set of images were obtained at $1 \mu m$ interval between $z = -5 \mu m$ and $z = 5 \mu m$, where the focused image is at $z = 0 \mu m$. The positively and negatively defocused images can be used to approximate the differential equation in equation (B.0.1). The phase of the modified regions were computed using QPM software (ITAIA) from the in-focus and defocused images.

$$\frac{\partial I(r)}{\partial z} = \frac{I_+ - I_-}{2z}, \quad (\text{B.0.2})$$

Assuming the modified structures are cylindrically symmetric about its axis, the phase $\phi(x, z)$ of the light passes through the structure is given by the Abel transform [152, 153]:

$$\phi(x, z) = \frac{4\pi}{\lambda} \int_x^R n(r, z) r (r^2 - x^2)^{-1/2} dr, \quad (\text{B.0.3})$$

where R is the radius of the structure, $n(r, z)$ is the refractive index, z is the

APPENDIX B. RECONSTRUCTING THE REFRACTIVE INDEX

direction along the fibre axis, r is the radial distance and x is the direction perpendicular to the light propagation.

Therefore refractive index of the structure can be obtained by obtaining the inverse Abel transform of the phase image [150, 152–154]:

$$n(r, z) = -\frac{\lambda}{2\pi^2} \int_0^R \frac{\partial\phi(x, z)}{\partial x} (R^2 - x^2)^{-1/2} dx. \quad (\text{B.0.4})$$

A number of algorithms have been developed over the years to obtain the refractive index $n(r, z)$ by solving either equation (B.0.3) or (B.0.4). In this work we use the Fourier decomposition based method proposed by Kalal and Nugent [152]. In the Fourier based method, the phase can be written as a Fourier cosine series as follows [150, 152]:

$$\phi(x, z) = a_0 + \sum_{m=1}^{\infty} a_m(y) \cos\left(\frac{m\pi x}{R}\right), \quad (\text{B.0.5})$$

where the a_m are Fourier coefficients, and the inverse Abel transform is applied to each Fourier components. Using the same notation, the transverse phase gradient is given by [150]:

$$\frac{\partial\phi(x, y)}{\partial x} = -\frac{\pi}{R} \sum_{m=1}^{\infty} m a_m(y) \sin\left(\frac{m\pi x}{R}\right). \quad (\text{B.0.6})$$

Hence by decomposing the transverse phase gradient $\frac{\partial\phi(x, y)}{\partial x}$ as Fourier series it is possible to determine the coefficients as $a_m(y)$ for $m \geq 1$ and the method of [152] was used to determine the refractive index profiles.

Figure B.2 shows in-focus and defocused images obtained at $\pm 2 \mu m$ from the in-focus image, of the structures inscribed by the *High-Rep 800 nm* femtosecond system. Images were obtained using a $20\times$ objective and a band pass filter centred at 531 nm with a bandwidth of 5 nm was used for filtering the while illuminations light. Figure B.3 shows the relative phase difference between the modified and unmodified regions and the reconstructed refractive index, location of the

APPENDIX B. RECONSTRUCTING THE REFRACTIVE INDEX

measurement of RI is indicated by blue line in the grey image, calculated using QPM software and reconstructed refractive index using inverse Abel transform.

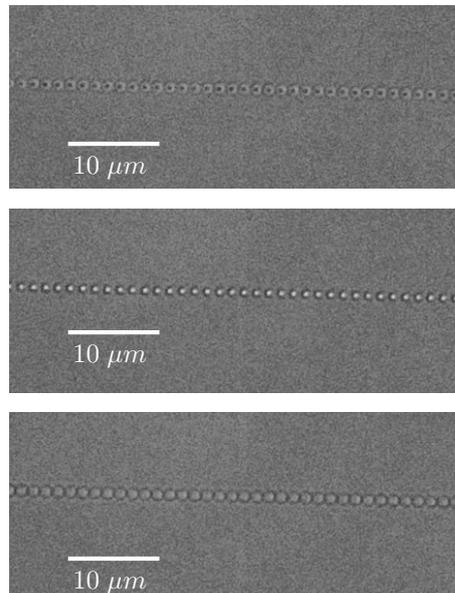


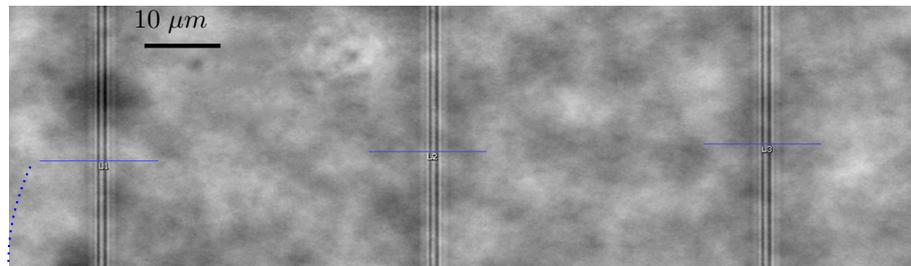
Figure B.2: Shows a typical example series of images taken in focus (middle), and $2 \mu\text{m}$ either side of the focus (top and bottom).

Alternatively, the refractive index of the modified regions can be approximated using the phase results as follows:

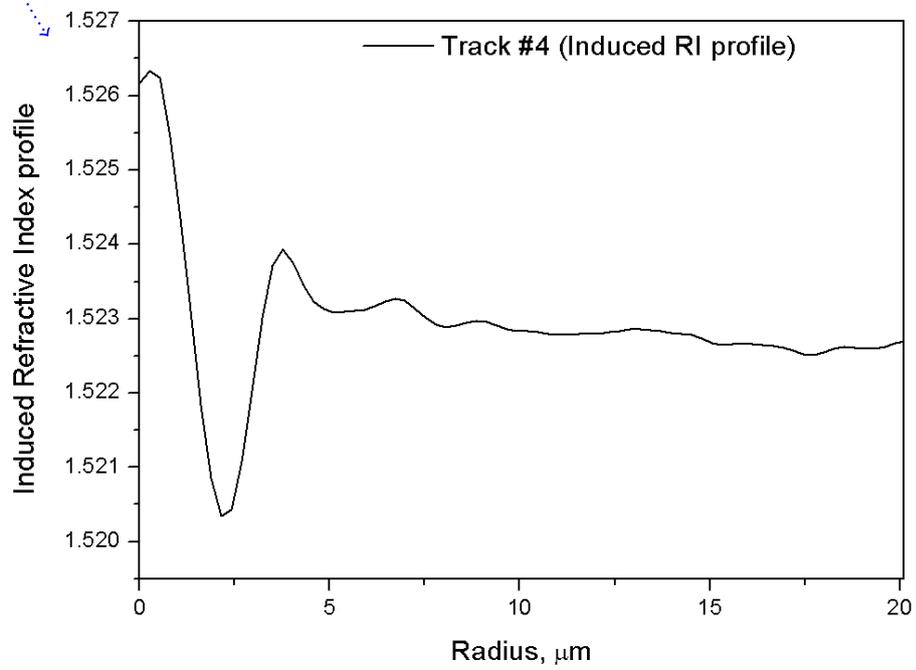
$$\Delta n = \frac{\Delta\phi\lambda}{2\pi d}, \quad (\text{B.0.7})$$

where λ is the wavelength of the filter used, d is the thickness of the modified regions, $\Delta\phi$ is the relative phase between unmodified and modified approximated by the QPM measurements.

APPENDIX B. RECONSTRUCTING THE REFRACTIVE INDEX



(a)



(b)

Figure B.3: A typical a) phase image and b) the reconstructed refractive index of the waveguides inscribed by High-Rep 800 nm .

List of publications

- [155] P. H. Tomlins, G. N. Smith, P. D. Woolliams, J. Rasakanthan, and K. Sugden, "Femtosecond laser micro-inscription of optical coherence tomography resolution test artifacts," *Biomed. Opt. Express*, vol. 2, pp. 1319–1327, May 2011.
- [156] J. Rasakanthan, K. Sugden, and P. H. Tomlins, "Processing and rendering of fourier domain optical coherence tomography images at a line rate over 524 khz using a graphics processing unit," *Journal of Biomedical Optics*, vol. 16, no. 2, p. 020505, 2011.
- [157] J. Rasakanthan, M. Baregheh, M. Dubov, V. Mezentsev, G. Smith, K. Sugden, P. D. Woolliams, and P. H. Tomlins, "Quantitative optical coherence tomography for characterization of microscopic structures with varying refractive index," in *CLEO/Europe and EQEC 2011 Conference Digest*, p. CH3 3, Optical Society of America, 2011.
- [158] J. Rasakanthan, G. C. B. Lee, P. D. Woolliams, and K. Sugden, "Parametric study of femtosecond inscription of microstructures for oct artefact fabrication," vol. 8213, p. 82133N, SPIE, 2012.
- [159] M. G. Sandrian, P. Tomlins, P. Woolliams, J. Rasakanthan, G. C. Lee, A. Yang, B. Povazay, A. Alex, K. Sugden, and W. Drexler, "Three-dimensional calibration targets for optical coherence tomography," vol. 8229, p. 822914, SPIE, 2012.
- [160] G. C. B. Lee, J. Rasakanthan, P. D. Woolliams, and K. Sugden, "Fabrication of high quality optical coherence tomography (oct) calibration artefacts using femtosecond inscription," vol. 8427, p. 84271K, SPIE, 2012.
- [161] K. Sugden, P. D. Woolliams, J. Rasakanthan, and P. Tomlins, "Femtosecond laser micro-inscription of phantoms for calibration and characterization of optical coherence tomography imaging systems," 2012.