UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS

School of Chemistry

Chemical Applications of *e*Science to Interfacial Spectroscopy

by

Jamie M. Robinson

A thesis for the degree of Doctor of Philosophy

May 2009

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING, SCIENCE AND MATHEMATICS
SCHOOL OF CHEMISTRY

Doctor of Philosophy

CHEMICAL APPLICATIONS OF ESCIENCE TO INTERFACIAL SPECTROSCOPY

by JAMIE M. ROBINSON

The UK *e*Science initiative has undertaken the unenviable task of providing the scientist with computational tools to enhance the way in which science is performed.

In this work, the experimental measurement of polarisation dependent reflectance second harmonic generation at the air/liquid interface was examined and tools produced to aid the process. All stages of the process are considered, the experimental stage, the data workup and publication of results.

Specifically covered were tools for remote monitoring and control of the experiment, automated archival and retrieval of raw data, and automated processing of data and handling of these derived results.

The tools were then used to investigate the behaviour of the laser dye Rhodamine 6G at the air/water interface. The ease and reliability of recall of data afforded by the tools caused concerns to be raised about the repeatability of the experiment, and hence the surface of pure water samples was also examined. The water data confirmed the acceptable bounds of the experimental results showing the Rhodamine 6G behaviour did not follow the accepted model for bulk concentration of around 1×10^{-5} mol dm^{-3} but did conform to the model at other concentrations.

Contents

1 Introduction **1**

2 Interfacial Chemistry **3**
2.1 Why study liquid interfaces? . . . 3
2.2 Surface tension . . . 3
2.3 Ellipsometry . . . 8
2.4 Non-linear optical techniques . . . 8

3 *e*Science **10**
3.1 What is *e*Science? . . . 10
3.2 The combinatorial approach . . . 10
3.3 Experiment monitoring . . . 11
3.4 Location awareness . . . 12
3.5 Remote control . . . 12
3.6 Data storage and accessibility . . . 13
3.7 "End-to-End" storage . . . 15
3.8 Publish@source, Publish in full . . . 16
3.9 Why use SHG as an example . . . 16

4 Laboratory Monitoring **18**
4.1 Background . . . 18
4.1.1 "SmartLabs" and Data Loggers . . . 18
4.1.2 Real-time data communication . . . 19
4.1.3 Publish/subscribe and broker technologies . . . 20
4.1.4 The hierarchical nature of broker topics . . . 21
4.1.5 Security for broker applications . . . 23
4.2 Implementation . . . 23
4.2.1 Implementation Overview . . . 23
4.2.2 Electronics . . . 24
4.2.3 MQTT messages in the SmartLab . . . 24
4.2.4 Transformation agents . . . 25
4.3 Discussion . . . 28

5 Location Awareness **30**
5.1 Background . . . 30
5.2 Implementation . . . 32
5.3 Ethical Implications . . . 32
5.4 Current Usage . . . 36

6 Remote Control **37**
6.1 Background . . . 37
6.2 Authorisation and access control . . . 38
6.3 Implementation . . . 39
6.3.1 LEGO construction . . . 40
6.3.2 RCX Software . . . 41
6.3.3 Microscope software . . . 42
6.3.4 Frontend control . . . 43
6.4 Discussion . . . 43

7 Second Harmonic Generation - The Chemistry Experiment **46**
7.1 Background theory . . . 46
7.1.1 Basic theory of SHG . . . 46
7.1.2 Nature of second harmonic generated emissions . . . 48
7.1.3 Origins of the liquid surface specificity of SHG . . . 48
7.1.4 Intensity of the second harmonic signal . . . 49
7.1.5 The effect of polarisation on second harmonic signal . . . 50
7.2 SHG optics . . . 51
7.2.1 Optics zero alignment . . . 52
7.3 Electronics . . . 53
7.4 Data collection . . . 53
7.5 Experimental procedures . . . 53
7.5.1 Glassware cleaning and sample makeup . . . 53
7.5.2 Apparatus Checks . . . 55
7.5.3 Compound behaviour . . . 56
7.5.4 Polarisation Scan . . . 58
7.5.5 Concentration Isotherm . . . 59

8 Second Harmonic Generation - The Data Collection Exercise **60**
8.1 Background . . . 60
8.1.1 The SHG experiment . . . 60
8.1.2 Data Storage . . . 61
8.1.3 Data Publication . . . 63
8.1.4 Data Review . . . 65
8.2 Implementation . . . 65
8.2.1 Data acquisition system . . . 65
8.2.2 Data Storage . . . 67
8.2.3 Publication @ Source . . . 70
8.2.4 Automatic publication . . . 71
8.2.5 Experimental data recall and analysis . . . 72
8.2.6 Grouping raw data into experiments . . . 75

9 Rhodamine 6G **81**
9.1 Previous Work . . . 82
9.1.1 UV-Visible Absorption spectroscopy . . . 82
9.1.2 Crystal Structure . . . 83
9.1.3 Surface Tension . . . 83

9.2 Experimental . . . 84
9.2.1 Ellipsometry . . . 84
9.2.2 Second Harmonic . . . 85
9.3 Results . . . 88
9.3.1 Surface Tension . . . 88
9.3.2 Ellipsometry . . . 89
9.3.3 SHG Polarisation results . . . 90
9.3.4 SHG Concentration Isotherm . . . 91
9.4 Discussion . . . 95
9.4.1 Surface Tension . . . 95
9.4.2 SHG Concentration Isotherm . . . 97
9.4.3 SHG Polarisation Dependence measurements . . . 99

10 Water **101**
10.1 Background . . . 101
10.2 Experimental . . . 103
10.3 Results . . . 106
10.4 Revisiting the R6G data . . . 111

11 The value of eScience to Interfacial Chemistry **113**

12 Future work **115**

A Glossary **119**
A.1 Chemistry Terms . . . 119
A.2 Computer Science Terms . . . 124

B Supporting Data CD **128**
B.1 CD Structure . . . 128
B.2 Software Manuals . . . 130
B.2.1 CombeChem MQTT Database Backup Agent . . . 131
B.2.2 CombeChem MQTT Mimic Agent . . . 133
B.2.3 CombeChem MQTT Agents - Linux Binary . . . 135
B.2.4 CombeChem laboratory monitoring recall tool - Web . . . 137
B.2.5 CombeChem MQTT NI-DAQ Windows capture agent . . . 138
B.2.6 CombeChem MQTT NI-DAQ Windows capture agent - BINARY . 140
B.2.7 CombeChem bluetooth device tracker . . . 142
B.2.8 CombeChem remote control tool . . . 144
B.2.9 CombeChem remote control modules . . . 145
B.2.10 CombeChem MQTT/Database Enabled SHG capture . . . 147
B.2.11 CombeChem SHG Experiment AutoBlogger . . . 148
B.2.12 CombeChem SHG Experiment database to review database, and automatic fit tools . . . 150
B.2.13 CombeChem SHG Web frontend to the experiment and review databases . . . 152

Publication List **153**

References **154**

List of Figures

2.1 Schematic showing the interface model used for the derivation. 4

4.1 Room layout of the simple lab and office example. 22
4.2 An example of the XML messages that carry sensor information in the SmartLab system. 25
4.3 Data-flows within the laboratory sensor and experiment messaging system. 26
4.4 Environment data recall showing the air-conditioner fault. 28
4.5 The mobile phone client 29

5.1 Example of the messages that carry location information. 32
5.2 Data flow within the bluetooth beacon client. 33
5.3 Data stored in the location database. 33
5.4 Data flow within the bluetooth database manager. 34

6.1 The IntelPlay microscope mounted in the LEGO frame. 40
6.2 Block diagram to represent the software interactions within the MQTT controlled LEGO microscope demonstrator. 41
6.3 Software flow within the RCX PC bridge agent. 42
6.4 Procedure flow within the RCX Code. 43
6.5 The end user interface to the microscope. 44

7.1 Optical arrangement used in this work. 52
7.2 Block diagram of the electronics in the SHG Rig. 54
7.3 Checking laser power stability. 56
7.4 Comparing Monochromator scans. 58
7.5 Process flow for an SHG experiment examining response with polarisation change. 59
7.6 Process flow for an SHG experiment examining one parameter. 59

8.1 A simple ‘triple’ relationship. 61
8.2 Triples to store a single measurement. 62
8.3 Extending the triples to further measurements. 62
8.4 Process flow within the experiment data capture software. 66
8.5 SHG Software - Login Screen. 68
8.6 SHG Software - Experiment Design. 68
8.7 SHG Software - Meta-Data Entry. 68
8.8 SHG Software - Data Collection. 68
8.9 Relational Database Scheme for storage of SHG related data. 69
8.10 The triscape browser. 71

8.11 Process flow within the auto-blog agent. 73
8.12 A typical SHG blog entry. 74
8.13 Overview of the experiment detail recall and extraction system. 75
8.14 SHG Review - Experiment Selection. 76
8.15 SHG Review - Experiment Display. 76
8.16 SHG Review - Experiment Comparison. 76
8.17 Relational Database Scheme for storage of SHG summary data. 77
8.18 Process used to group raw data into experiments. 78
8.19 Process used to translate a list of experiment ID's into a human readable summary. 79
8.20 Process flow for the automatic fitting of polarisation scanning SHG experiments to an existing model for storage with the summary. 80

9.1 The molecular structure of Rhodamine 6G. 81
9.2 Published UV-Vis spectra for R6G in ethanol solution. 82
9.3 Crystal structure of R6G. 83
9.4 Surface Tension of Aqueous R6G solution. 84
9.5 SHG signal power dependence for R6G(aq) solutions. 86
9.6 Monochromator scans for a 10^{-7} mol dm^{-}3 R6G solution. 87
9.7 Monochromator scans for 10^{-5} mol dm^{-3} R6G solution. 87
9.8 Surface Tension of Aqueous R6G solution. 88
9.9 Refractive Index derived from ellipsometric measurements of R6G. 90
9.10 R6G polarisation scans at various bulk concentrations. 91
9.11 R6G polarisation scans for bulk concentrations 1×10^{-5} mol dm^{-3} and 5×10^{-5} mol dm^{-3}. 92
9.12 Raw data from a concentration dependent SHG experiment. 93
9.13 Scaled data from a concentration dependant SHG experiment. 94
9.14 Combining many concentration dependant SHG experiments improves confidence in the data. 95
9.15 Surface Tension of Aqueous R6G solution. 96
9.16 Analysing the concentration dependant experiment. 98

10.1 S_{out} SHG polarisation dependence for achiral samples. 102
10.2 Sample monochromator scan for water. 104
10.3 Sample power dependence of water SHG. 104
10.4 An example of a "repeated polarisation" experiment. 105
10.5 Sample fitted polarisation plot for a recent water sample. 106
10.6 Box-plot of the all-real water ratios. 111

A.1 Schematic of light reflecting and refracting from an interface. 121

List of Tables

8.1 Separating the SHG data fields according to their frequency of change. . . 70

9.1 Surface Tension of Aqueous R6G solutions. 89
9.2 Surface Tension of Aqueous R6G solutions. 97
9.3 Model Co-efficients for R6G polarisation SHG data. 99

10.1 Achiral S_{out} diagnostics table. 103
10.2 ABC values for water SHG polarisation dependence for source 1. 107
10.3 ABC values for water SHG polarisation dependence. 108
10.4 Water SHG polarisation fit ratios. 109
10.5 ABC values for water SHG polarisation dependence, A,B and C assumed to be real. 110
10.6 Water SHG polarisation fit ratio summary. 111

Acknowledgements

The author would like to thank Prof. Jeremy Frey, Prof. John Dyke, Prof. David De Roure, and Dr Andy Stanford-Clark for their advice and academic input, The research groups of Frey, De Roure and Standford-Clark for their assistance with practical matters and for sample code, and his house-mates Adrian and Kieran for their support during preparation of this document.

Chapter 1

Introduction

In this work the experimental measurement of polarisation dependent reflectance second harmonic generation (SHG) at the air liquid interface is examined, and tools provided to aid the process. All stages of the procedure are investigated; the experimental stage, the data workup, and the publication of results.

The study of interfaces has received increasing attention in recent years, in part due to the abundance of interfaces in everyday processes, and in part through technological advances. Surface second harmonic generation (SHG) is an example of a high intensity non-linear technique, made possible by the invention of the laser in the 1960's. The SHG apparatus used in this work was developed as part of ongoing research.

One of the aims of the Comb*e*chem project (of which this work is a part) was to provide an "end-to-end" publishing mechanism, following a molecule from synthesis to scientific publication. An important step in this chain is characterisation of the substance. For the purpose of a demonstrator, compound characterisation could be covered by implementing a simple database storage system for a refined technique such as NMR where the results are reliable. However another goal of comb*e*chem is to provide tools to allow scientists (chemists in particular) to work more effectively in the laboratory. Part of this is enabling them to make informed decisions about the quality of data collected. To this end the experiment being used needs to be easily accessible to allow the inclusion of additional sensors to monitor laboratory conditions. This is somewhat easier for an experiment that is still under development, such as the SHG experiment at Southampton. The SHG experiment uses stock electronics and purpose-built control software, hence modifications to facilitate "SmartLab" techniques are somewhat easier to implement than they may be with a commercially built, closed source system.

The analysis of SHG data is a reasonably involved procedure, during which a number of assumptions have to be made. Altering these assumptions may lead to different results, hence it is important to record the exact steps performed in making a particular analysis. This fits well with the "publish@source" ideas that comb*e*chem set out to achieve.

Specifically covered by this work are tools for remote monitoring of the experiment, automated archival and pre-publication of the raw data, and the design of architecture to allow the capture and publication of the data workup procedure that leads to a published work on the topic. These tools are then used to assist in a study of the laser dye Rhodamine 6G (R6G) using SHG. The results from a single technique may be inconclusive for the study of a system, hence, as is demonstrated here, it is often useful to combine the outcomes from a range of techniques when drawing a conclusion.

Structure of this thesis

The thesis is structured such that the reader is led through the topics covered, starting with an introduction to Interfacial Chemistry techniques in chapter 2 and *e*Science in chapter 3. Discussion then covers the main threads of this work individually, starting with Laboratory Monitoring (Chapter 4), Location Awareness (Chapter 5) and Remote Control of Laboratory apparatus (Chapter 6) before moving on to examine the technique of Second Harmonic Generation, both as a measurement technique (Chapter 7) and as an example of the *e*Science Techniques discussed (Chapter 8). Finally the study of Rhodamine 6G using the technology developed is presented in Chapter 9. This study proved inconclusive, so the simpler system of pure water is considered in Chapter 10. Possible extensions of this work into other projects is explored in Chapter 12. A glossary is provided in Appendix A which aims to explain some of the terminology used in this thesis from both the computer science and chemistry perspectives.

Chapter 2

Interfacial Chemistry

2.1 Why study liquid interfaces?

Liquid interfaces play a fundamental role in science and technology; for example, the air/liquid and liquid/liquid interfaces present in biological systems, environmental processes and surfactant chemistry. The development of surface-sensitive probes capable of yielding molecular information about these interfaces is of great importance to aid the understanding of them and their role in complex systems.

2.2 Surface tension

At surfaces the asymmetry of the system results in the generation of surface tension. This is the result of the molecules at the surface seeking to be in the lowest energy situation, the bulk, and hence rearranging to minimise the surface area.

The amount of work required to change the surface area of a sample by some infinitessimal amount dA is proportional to magnitude of that area change. This can be expressed as equation 2.1[1].

$$dw = \gamma dA \tag{2.1}$$

Where γ is a proportionality constant, commonly named the surface tension which has units of J m^{-2}, although it is more common to find γ expressed in units of N m^{-1}. The surface tension of many liquids has been measured, and deviations from these standard values can be an indication of impurities. However, surface tension is a macroscopic measurement which yields no information about the nature of the adsorbed species. Gibbs showed that surface tension is related to adsorption of the components of a system as follows:

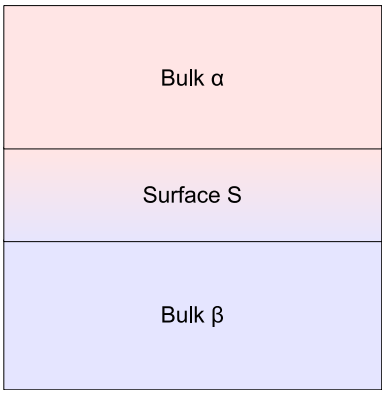


FIGURE 2.1: Schematic showing the interface model used here.

It is assumed that a generic surface can be represented as 3 layers (as depicted in figure 2.1). The internal energy, entropy and Gibbs energy of the system can then be considered as the sum of their components in the three layers:

$$U = U^{\alpha} + U^{\beta} + U^{s} \tag{2.2}$$

$$S = S^{\alpha} + S^{\beta} + S^{s} \tag{2.3}$$

$$G = G^{\alpha} + G^{\beta} + G^{s} \tag{2.4}$$

Similarly the quantity of solute, n, can be considered as:

$$n = n^{\alpha} + n^{\beta} + n^{s} \tag{2.5}$$

If we recall the definition of Gibbs energy in terms of enthalpy and entropy[2]:

$$G = H - TS \tag{2.6}$$

And that enthalpy can be defined in terms of internal energy and physical parameters[2]:

$$H = U + pV \tag{2.7}$$

Then we can write relations for infinitesimal changes in Gibbs energy under conditions of constant temperature and pressure[2]:

$$\left(\frac{\partial G}{\partial T}\right)_{p} = -S \tag{2.8}$$

$$\left(\frac{\partial G}{\partial p}\right)_{T} = V \tag{2.9}$$

Similarly if we recall the definition of Surface Tension in equation 2.1 and that under

conditions of constant temperature and pressure the work term δw can be equated to Gibbs energy[2], it can be written:

$$\left(\frac{\partial G}{\partial A}\right)_{T,p,n_j} = \gamma \tag{2.10}$$

where A is surface area.

The chemical potential is defined as the rate of change of Gibbs energy with quantity at constant temperature, pressure and surface area[3]:

$$\mu_i = \left(\frac{\partial G}{\partial n_i}\right)_{T,p,A,n_j} \tag{2.11}$$

which is the definition of μ_i for a multicomponent system. n_j is shorthand notation for $n_{j\neq i}$.

Hence by considering the combination of these, changes in G at the surface can be written as[4]:

$$dG = Vdp - SdT + \gamma dA + \sum \mu_i dn_i \tag{2.12}$$

Given that these experiments are usually performed at room temperature and pressure, these can be considered constant, hence we need only consider the extensive variables:

$$dG = \gamma dA + \sum \mu_i dn_i \tag{2.13}$$

Integrating this, under conditions of constant γ and μ_i gives the total gibbs energy at constant temperature and pressure:

$$G = \gamma A + \sum \mu_i n_i \tag{2.14}$$

Comparing equations 2.13 and 2.14 yields the Gibbs-Duhem equation

$$0 = Ad\gamma + \sum n_i d\mu_i \tag{2.15}$$

from which the following can be written

$$d\gamma = -\sum \frac{n_i}{A} d\mu_i \tag{2.16}$$

If we define the surface concentration (also known as the adsorption of the surface) as the amount of material per unit area:

$$\Gamma_i = \frac{n_i}{A} \tag{2.17}$$

$$\therefore \Gamma_i = -\left(\frac{d\gamma}{d\mu_i}\right)_{T,p,\mu_j} \tag{2.18}$$

where Γ is the adsorption of the species at the interface (in units of surface concentration) and μ is the chemical potential.

If we consider a surface onto which molecules are being adsorbed and desorbed, then the rate of adsorption will be proportional to the concentration and coverage of the surface, hence could be written as

$$R_{ads} = k_a c(1 - \Theta) \tag{2.19}$$

where k_a is the adsorption rate constant, c is the concentration and Θ is the surface coverage.

Similarly, the rate of desorption is proportional to the surface coverage, hence

$$R_{des} = k_d \Theta \tag{2.20}$$

When the system is at equilibrium, the rate of adsorption and desorption will be equal, hence combining 2.19 and 2.20:

$$k_a c(1 - \Theta) = k_d \Theta \tag{2.21}$$

$$\Rightarrow \frac{\Theta}{1 - \Theta} = \frac{k_a}{k_d} c = Kc \tag{2.22}$$

where K is the system Langmuir adsorption constant.

$$\Rightarrow \frac{1 - \Theta}{\Theta} = \frac{1}{Kc} = \frac{1}{\Theta} - 1 \tag{2.23}$$

$$\Rightarrow \frac{1}{\Theta} = 1 + \frac{1}{Kc} = \frac{1 + Kc}{Kc} \tag{2.24}$$

$$\Rightarrow \Theta = \frac{Kc}{1 + Kc} \tag{2.25}$$

The surface coverage, Θ, is the fraction of the surface which is covered by the sample molecules, hence can also be expressed as the ratio of the adsorption of the sample over

the maximum theoretical adsorption.

$$\Theta_i = \frac{\Gamma_i}{\Gamma_\infty} \tag{2.26}$$

Hence, for a single adsorbate, we state that

$$\Theta = \frac{\Gamma}{\Gamma_\infty} = \frac{Kc}{1+Kc} \tag{2.27}$$

$$\Rightarrow \Gamma = \frac{\Gamma_\infty Kc}{1+Kc} \tag{2.28}$$

If we define chemical potential in terms of activity for a solute-based standard state:

$$\mu_i = \mu_i^{\ominus} + RT \ln a_i = \mu_i^{\ominus} + RT \ln(f_i c_i) \tag{2.29}$$

where f_i is the activity coefficient. Then assume that the system follows Henry's law, i.e. that it tends to ideal behaviour as c_i tends to 0, hence under ideal conditions $c_i \to 0$, $f_i \to 1$ therefore it can be written,

$$d\mu_i = RT d \ln c_i \tag{2.30}$$

combining this with equation 2.18 then yields

$$RT\Gamma_i = -\left(\frac{d\gamma}{dlnc_i}\right)_{T,p,\mu_j} \tag{2.31}$$

where j signifies solvent, therefore μ_j is constant under ideal, dilute, conditions. Combining this with 2.28 it can be written

$$\left(\frac{d\gamma}{dlnc_i}\right)_{T,p,\mu_j} = -RT\frac{\Gamma_\infty Kc_i}{1+Kc_i} \tag{2.32}$$

By writing

$$\frac{d\gamma}{dlnc} = \frac{d\gamma}{dc}.\frac{dc}{dlnc} = c\frac{d\gamma}{dc} \tag{2.33}$$

2.32 can be simplified to

$$-\left(\frac{d\gamma}{dc_i}\right)_{T,p,\mu_j} = \frac{RT\Gamma_\infty K}{1+Kc_i} \tag{2.34}$$

hence by integrating,

$$\gamma = C + RT\Gamma_\infty ln(1+Kc_i) \tag{2.35}$$

where C is a constant.

By inspection it can be seen that in the case of pure solvent ($c_i = 0$) the constant must be equal to the surface tension of the pure solvent for this to hold true, and if we consider a single solute $c_i = c_s$, hence for aqueous, single solute, systems we can write

$$\gamma = \gamma_w + RT\Gamma_\infty ln(1 + Kc_s) \tag{2.36}$$

which can be fitted to experimental data using a regression technique, yielding numerical values for Γ_∞ and K, which can be used with equation 2.27 to yield surface coverage for a given concentration.

2.3 Ellipsometry

Ellipsometry is an optical technique for the study of thin films, which exploits the polarisation transformation that occurs when a beam of light is reflected from an interface. It is able to characterise the interface, and to observe events that occur there.

A beam of light of controlled polarisation is reflected from the sample's surface and by monitoring the changes in the polarisation of the light, information about the dielectric properties of the layer can be deduced. In some cases it is possible to yield information down to the thickness of a single atomic layer. Properties which may be ascertained by this technique include layer thickness and a range of optical constants. By probing the complex refractive index of a solid layer, fundamental physical parameters such as crystal quality, chemical composition and electrical conductivity can be deduced.[5] Ellipsometry probes the surface using optical radiation, hence is able to operate under ambient conditions unlike many other (solid) surface probes which can only operate under vacuum conditions through reliance on subatomic particles as probes. The ability to operate ellipsometry under ambient conditions allows for it to probe liquid surfaces, although these provide more technological challenges than solids. It is, however, possible to gain information about many surface properties, including refractive index and layer thickness.

In this work, ellipsometry is used as a tool to determine the complex refractive indexes required for analysis of the SHG data. It also acts as an additional indicator to surface trends that are dependent on concentration.

2.4 Non-linear optical techniques

Nonlinear optics is the study of phenomena that occur when a material is subjected to a high power light source which causes a change in its optical properties. Currently, only

laser sources are sufficiently intense to cause these phenomena. More generally nonlinear optical phenomena are those whose response is nonlinear with respect to the intensity of the applied light source. The earliest example of this is the discovery of second harmonic generation by Franken et al. in 1961[6]. The second harmonic response was found to be dependant on the square of the intensity of the incident light; it is a second order non-linear optical effect.

Sum Frequency Generation (SFG) and Second Harmonic Generation (SHG)

Two important non-linear techniques are those of second harmonic generation and sum frequency generation[7, 8, 9].

In the sum frequency generation experiment, a sample in subjected to two light sources of differing frequencies ($\omega_1$and ω_2). The sample then emits radiation at the sum of these frequencies ($\omega = \omega_1 + \omega_2$). Second harmonic generation (SHG) is the degenerate case when $\omega_1 = \omega_2$, this is typically written as input frequency ω producing radiation at 2ω.

When considering fluid systems, second order non-linear optical techniques are inherently surface specific. The electric dipole approximation shows that in centrosymmetric media (i.e. the bulk) second order non-linear effects are electric dipole forbidden[7]. Second harmonic generation is therefore a useful probe for the investigation of buried interfaces as well as exposed ones (assuming that at least one of the bulk media is transparent to the light at ω and 2ω). This high surface specificity also allows for monitoring of structure and properties down to sub-monolayer coverage levels.

A common application of Second harmonic generation is in the output stage of a frequency doubled laser, an example of this is the "green" Nd:YAG laser; Nd:YAG lases at 1064 nm, the output from this is frequency doubled within the housing of the laser head to produce the 532 nm light that the user receives. Second harmonic generation is also useful for the study of liquid interfaces, as demonstrated here, and also more recently in biological applications, such Campagnola and Loew's work[10] studying labelled biological molecules. Second harmonic generation is discussed in more detail in chapter 7.

Chapter 3

*e*Science

3.1 What is *e*Science?

"*e*Science refers to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet."[11]

Broadly speaking *e*Science is the use of computers to enhance science. The UK *e*Science programme covered a vast range of topics including Clinical Science support (CLEF)[12], Biomolecular Simulation (BiosimGRID)[13], Distributed Aircraft Maintenance (DAME)[14], Collaborative Knowledge (CoAKTing)[15] and providing support methodologies; such as storage and resource management (MyGRID)[16]. The work presented here has been within the Comb*e*chem[17] project; Comb*e*chem was investigating end-to-end data storage for the chemical process. This ranges from electronic laboratory notebooks in the organic synthesis lab[18] through automated crystal structure collection, resolution and publication[19], data collection for laser driven surface spectroscopy[20] to *ab-initio* chemical property calculation[21], and the use of statistical methods to improve laboratory efficiency[22].

3.2 The combinatorial approach

In chemical science there is an increasing trend to use high throughput methodologies and technologies, both in the parallel production of libraries of compounds, for example for drug discovery applications, and also in the processing and characterisation of these molecules by spectroscopic techniques. This has been aided both by automation technologies allowing instruments to run for longer without user intervention, a "dark-lab" scenario, and by the increasing pervasiveness of instrumentation. Instruments which were previously expensive and required specialist operators, are now becoming available

as off-the-shelf low maintenance systems. Equally, other experimental techniques are becoming available and so providing a wider range of potential data sources.

This ever increasing range of compounds and results presents a challenge to the scientific community. The traditional method of peer review publishing has already exceeded its capacity, resulting in a large amount of chemical information remaining unpublished within the academic community and often forgotten or mislaid when it doesn't quite fit with the latest "hot topic" or funding proposal. By not publishing this information it is effectively lost; inaccessible to people outside the initial institution and often forgotten by the scientists within it. By storing such information, as its collected, in a well indexed system the results became locatable and then, when required, retrieval is a far easier task. Indeed, it would even be possible for other scientists within the institution and beyond to find the information and gain access to it - subject to having adequate permissions, of course. A discussion of potential authorisation techniques is included in section 6.2 as part of the examination of remote control techniques.

There is a danger that such a system could become flooded with poor quality, unreliable data. To aid the user in distinguishing this, a number of things can be done. Firstly if all the raw data making up a final result and appropriate provenance information for the systems being studied are published in a transparent manner, then user confidence can be increased. Secondly, by combining a number of related data sources from a range of researchers, correlations may be revealed. This will allow conclusions to be drawn when previously they would not; for example, by corroborating a previously inconclusive result. Conversely, if contradicting results are shown then confidence in a data-set may be reduced. This could prompt further experiments to confirm the correct conclusion and avoid the premature acceptance of incorrect conclusions.

In summary, a system of trust needs to be built up. In the traditional publishing methodology, this trust is based on the review of papers by academic peers; by the trust in their expertise that they will be able to identify faulty conclusions. The suggestion isn't that this should not occur, but that the reviewers judgement can be augmented by the author providing transparent access to their raw data, and presenting proof of the provenance of their data and techniques.

3.3 Experiment monitoring

As experiments become more autonomous, the scientist is freed to perform other tasks during data collection. However it is important to periodically check the state of data capture. The most obvious way to do this is to physically view the experiment. It may not always be safe or sensible to do this; being present in the laboratory may affect the results, or part of the technique may pose a significant health hazard; for example, a liquid surface may be disrupted by the vibrations caused by people moving in the lab, and

an experiment containing an ionising radiation source may present a significant hazard to the user. In these cases the ability to remotely view the experiment is useful. With a computer controlled experiment, it may be adequate to use a remote screen viewing application (to view the apparatus control computers output). However these system do not scale well, and they often provide unnecessary functionality, such as remote control. In these cases it may be more appropriate to relay the most recent data set and current status to the user.

One should also consider environmental factors. Some experimental techniques can be very sensitive to changes in temperature, humidity, ambient light level, or be affected by vibrations. The importance of keeping a record of such conditions alongside experimental data should not be overlooked. It is often reasonable to configure the environmental monitors to run independently from the experiment, hence allowing monitoring of the space at all times. One needs to be able to inspect for correlations between experiment and environment, hence a way of combining these two data sources must also be considered.

3.4 Location awareness

When providing the user with feedback there are a number of competing technologies available; each providing a different level of detail, and possibility for interaction. Some of these may be location dependant, for example a physical dial on a rig, or a fixed computer screen. If the system was able to know the (approximate) location of the scientist, then it can provide feedback using the most appropriate display available.

For a number of years, researchers in the semantic computing field have recognised the value of location awareness, and a number of systems for tracking individuals have evolved. These are typically based around the user carrying a remote readable, unique identity tag that the system can identify and link to that individual, or that allow the user to know their location based on broadcasts from nearby beacons[23].

3.5 Remote control

There are situations where, in addition to the remote monitoring, it would be advantageous to be able to have remote input into the running of the experiment, this could be simple interactions such as authorising a long data collection based on a smaller preview set, or more complex interactions, for example, a complete reconfiguration of the experimental parameters. Technology suitable to the situations encountered must be put into place for such a system to be acceptable and where this can be integrated with existing infrastructure there are obvious savings to be made. Care must be taken to ensure that

the systems are incapable of placing personnel in dangerous situations and are resilient to failures caused by faulty data or malicious users. The dangers of this were highlighted by early computer controlled radio-therapy machines in which it was assumed that the computer control system was infallible and the instrument mechanics totally reliable. When these control systems operated outside their expected parameters an unacceptably high radiation dose was delivered to the patient, often with tragic results[24, 25]. Such accidents can be avoided using a combination of monitoring and hardware interlocks. When dealing with safety critical devices, such as laser interlocks, it is prudent to make them closed loop, and independent of networked messaging systems, the delay in a message being transferred may make the significant difference between injury and death. But providing read-only monitoring of these systems can be very useful, as the messaging system can then alert other people that an interlock event has occurred and that intervention is needed. Where a computer control system needs to interface with safety-critical equipment, it must be done in such a way that the system 'fails safe', i.e. that a failure results in as benign and stable a state as possible, and that warnings and alarms default to the "on" state. Computer activation of a system should only honoured when all the mechanical and electronic interlocks have given the "all clear".

This type of control is demonstrated by the control systems in entertainment display lasers. These systems allow a laser beam to be positioned by scanning mirrors, drawing shapes in the air over the audience. For one of these systems to be awarded a safety certificate it must demonstrate that all the mirrors position the beam into an internal beam-stop should any part of the system fail, where failures include power and controller. Often these fail-safes are mechanical, relying on gravity or springs.

3.6 Data storage and accessibility

To support the quantity of data that current scientific techniques can produce, efficient storage solutions must be found so that data loss can be avoided. With current low throughput methods, most chemists rely on hand written lab-books with spectra and results either being stuck in or filed separately with some identification system. By providing centralised data resources these spectra can be stored in a reliable, retrievable manner; retrievable not only by the chemist, but where appropriate, by their colleagues and the community at large. In terms of data throughput, a synthetic chemist may produce a couple A4 sides of written notes, and 3 or 4 printed spectra for a simple preparative sequence, although this may be more for a complex or difficult compound (where, for example, multiple purification steps, and purity checks may be required). The majority of spectra will be at most a few hundred kilobytes, so a rough estimation indicates that each preparation will produce about 500 kB of data over the space of about a week. It is important to bear in mind, however, that most chemists will produce a number of compounds in parallel. Furthermore, few researchers will be content with

just producing chemicals and many will then want to spend weeks investigating a few specialist properties; producing anything from a few lines of apparatus readings in a log book to many hundreds of megabytes of data.

Why stop at spectra? By extending the data storage systems to include the lab book as well, end to end data management can be achieved. By transferring the lab book to an electronic storage system, one can also help to protect against data loss. In the traditional paper book methodology there is typically a single point of failure. For example it is common (and useful) to have the lab book on the lab bench whilst performing preparations, however something as simple as a solvent spillage could render the content of the book useless. In an electronic system the same solvent spillage may render the terminal useless, but all data would be stored safely on a central server, ready to be recalled via another terminal. It is this security of data which has most impressed practical chemists with this type of scheme. Care must be taken in the design and installation of such infrastructure, to ensure that the back-end storage is adequately protected against data loss, through either hardware failure, building failure, or human error; in the case of a solvent spillage onto a notebook, an institution may loose a few weeks of one researchers work, but given the small amounts of data being handled, and the low cost of large amounts of storage, the failure of a server hard disk could loose the research output of a group, or worse department, for a much larger time frame. The low cost of storage is however the saving grace in this matter in that the financial excuse for not keeping adequate backups is no longer valid.

Consideration must also be given to the way in which users interact with the system, a recent study produced the quote "I can go anywhere and it's, like, this is me and my data. It's all there!"[26] which related a chemists enthusiasm towards the data security that could be afforded by a remote data storage facility. However, data storage isn't the end of the story for the electronic lab-book. An electronic system must be at least as useful as its traditional counterpart. But how to define useful? We can start with Dix's notion of used: "If a system is not used it is useless"[27]. This sounds rather obvious, however the best definition of a system's usability comes from the end user, and whether they will drop a current system in preference for a new one. Previous attempts to produce smart lab-books, both from commercial [28, 29, 30, 31] and research projects [32, 33, 34] have led to systems that are poorly regarded within the community[26]. In many cases this can be attributed to the interfacing of the system. They are generally based on the user's desktop PC, which does not translate well into the research laboratory where there is rarely the free space for computer systems, unless they are directly linked to the experiment. The geographic specificity of a single PC will also hinder this, as can the cost of providing multiple systems were space to allow this.

In their study, schraefel *et al*[26] attributed the success of their system to an understanding of how the chemist uses a lab book through an analogy based design process. They recognised the "hug-ability" of the paper lab-book, and embodied this in a wirelessly

networked tablet PC. This has the benefits of being portable (it is of similar size to an A4 notebook), allows for "freeform" sketching and note taking, and has remote storage benefits, by being always "online" and connected to the back-end database through a wireless network link. By regularly committing data to the data store, the chemist is able to recover from both software and hardware failure with minimal data loss.

3.7 "End-to-End" storage

By placing the record of the chemical process into an electronic storage facility from its inception, the route is opened for completing the end to end system. Taking the example of an organic synthesis, if at the planning stage the sample is given a unique identifier, or URI (Uniform Resource Identifier), this value can travel with the sample throughout its life-cycle. When a spectroscopic technique is performed on the sample, by using its URI, the spectra could be automatically linked to the record of that sample. When a material is taken on as the precursor to another, by using its URI a provenance chain is created. These lab samples would be associated with an individual, who would be a member of a research group. Therefore if that researcher's supervisor needs to be consulted about a problem, a chain of association could be set up (Chemist is member of group, Supervisor leads group) and hence the supervisor could access the spectroscopic data, along with the experimental observations, securely. A more in depth discussion of some of the technologies that can be used to implement this is presented in section 6.2

This is even more important when a group member leaves (or worse is suddenly removed - the "run down by a bus" problem), as at best experiment indexing currently occurs in a highly personal way, using dates or cryptic codes, or more often doesn't occur at all. The bringing together of data in this way also helps to avoid the situation where one may have a chemists lab-book, but have no idea which of the computer systems or filing cabinets stores their supporting data. All data in the proposed system will have clear identifiers, and ideally the data will be directly recallable.

When recording experimental processes, a trade off is usually made between verbosity and time. As a technique becomes familiar it is common for the amount of supporting data recorded to reduce; possibly to a level when the usefulness of the results becomes questionable. When recording experiment details it is easy to assume that a set of "standard" conditions and apparatus configurations was used. However, when revisiting these data later, trying to trace what the "standard" conditions were at the time of experiment becomes increasingly difficult. It is often impossible to trace such conditions in someone else's work. When using an electronic interface to record an experiment, prompts to record such information can be added or, if funds allow, some apparatus is now available which allows for the data logging systems to record the current setup.

Enforcing the recording of meta-data in this way improves the re-usability of information and adds to the provenance trail.

3.8 Publish@source, Publish in full

This whole package of data storage throughout the experiment leads the way for producing better publications. Making information about experimental conditions available should allow for other scientists to reproduce work more accurately than by trying to follow the cryptic methods found in the current literature.

In 2004, Frey and Hursthouse[35] noted that the current publishing methods, whilst being an important storage and archival mechanism (paper doesn't require special hardware to view), are unable to contend with the speed at which new science is being performed. They took the example of X-Ray Crystallography; at the time the published data set of structures was in the region of 300,000 which is collated in the Cambridge Structures Database[36], however they estimated that approximately 1.5 million structures have been successfully resolved out of a potential data set of 10 million compounds. The disparity between the 1.5 million structures known, and the 300,000 published was their major point of concern. They proposed and implemented mechanisms for self publication of data that allow all data to be published, and by providing a traceable provenance for the material right back to its preparation, they propose that end users will be better placed to make judgements about the quality of the data. Part of this provenance chain will be access to all the raw data from the analysis techniques carried out on a compound. This will also allow for the end user to reprocess the data should there be doubt in the technique used, or if a new technique becomes available.

3.9 Why use SHG as an example

As highlighted in the introduction the comb*e*chem project sought to provide the chemist with the tools necessary to reach an informed decision about the quality of experimental work. Improvements in automation technology have made it increasingly common to leave practical experiments in chemistry running unattended. In many cases this can lead to a safer working environment (for example, experiments involving ionising radiation or laser sources) and better results (for example, liquid surface experiments can be sensitive to vibration). However, by being present during an experiment an experienced chemist will notice problems as they occur and either alleviate them or abort the experiment early and restart it, having taken steps to avoid the problem recurring. Hence the chemist needs to be provided with the ability to monitor the experiment and its environment whilst taking advantage of the safety and quality of result improvements that can be gained by allowing an experiment to run unattended. It is also important to be able to

review these external factors when analysing data, as they may provide indicators that explain poor quality or unexpected results.

An experiment such as the collection of SHG data from liquid surfaces provides an example where a number of external, but monitorable, factors influence the results collected. It includes a high power laser source, making automated operation desirable, and a sample that is very sensitive to vibrations. In the Southampton experiment, light reflected from the surface is collected through a 1 millimetre monochromator slit, 1 metre away from the slit, hence the smallest of deflections at the sample surface can cause the beam to completely miss the detector.

Chapter 4

Laboratory Monitoring

4.1 Background

4.1.1 "SmartLabs" and Data Loggers

Improvements in data handling techniques have allowed for larger quantities of data to be managed sensibly. Considering the improvements in compound production methods mentioned in section 3.2, it is logical to conclude that spectroscopists will look for ways to improve throughput of the techniques used to characterise them. The current preference is to move towards automated systems, both to allow an individual experiment to run unassisted and also in the form of sample changing robots, allowing multiple samples to run without intervention. The introduction of automation can lead to improvements in the reliability and reproducibility of results as random human errors are removed (or at least significantly reduced). However we must remember that these systems are not infallible, and results which appear erroneous should be questioned. It is likely that experimental throughput will soon reach a level which renders human checking of individual spectra infeasible. In these cases, the computer systems should be equipped to identify errors. Where a well understood theory can lead to an in-silico prediction of the result, then this could be checked against the recorded data, highlighting problematic data for human verification. In a truly automated system, however, a computer agent would conduct such identification and schedule samples for further analysis where results prove to be substandard.

It has already been discussed that human factors can lead to anomalous data, but environmental factors should also be considered. Some experimental techniques can be very sensitive to the effects of such parameters as temperature, humidity, ambient light level, and vibration. The importance of keeping a record of such conditions alongside experimental data should not be overlooked. It is increasingly common to install systems in

laboratories to control these factors, for instance, air conditioners. However accompanying this is the risk of complacency, as one may start to rely on such systems, and assume that they function flawlessly. Part of a laboratory monitoring system should be keeping check, such that failure of these systems can be avoided or at least detected and an alert raised. Environmental monitoring would usually be installed independently from the experiment, this allows for checking of conditions outside the experimental time-frame; however one needs to be able to inspect data and records for correlations between experiment and environment. If the laboratory's experimental and environmental data acquisition systems are both configured to run from a single time-source, then data from these diverse sources can be correlated based on time stamps attached to each measurement.

4.1.2 Real-time data communication

As experiments become more autonomous, the scientist is freed to perform other tasks during data collection. It is important, however, for him to periodically check the status of data capture during an experimental run. The most obvious way to do this is to physically view the experiment, but it may not always be safe or sensible to do this as being present in the lab may affect results. In this case, the ability to remotely view the experiment is useful. With a computer controlled experiment, it may be adequate to use a remote screen viewing application such as VNC[37]. VNC allows a computer system to be controlled remotely over a TCP/IP LAN or the Internet. A window on the local machine is opened, which mimics the screen output of the remote machine. Local key presses and mouse gestures are transmitted back to the remote machine. However, this system does not scale well as there is a maximum of 1 client per remote machine.

It will often be the case that when interaction with software is required, physical interaction with the apparatus is also needed. In these cases it is more appropriate to relay the most recent data set and current status (for example "collecting data", "moving motors", or "waiting"). This runs hand in hand with the concept of publishing data into a database. If data is published periodically throughout the experiment, then the monitor client can recall this data for display purposes. By using the database as a proxy in this way, some attractive security bonuses can also be gained. The data could even be displayed in a web-page (most web-server scripting languages can interface directly with databases) giving, possibly, the ultimate in portability, as well written web pages can be viewed by the vast majority of current computer systems, potentially from anywhere in the world.

Automated systems will often need periodic human intervention. In order to maximise productivity, the system must be able to recognise when this interaction is required, and summon appropriate assistance. There currently exist a wide range of options for contacting people, ranging from dedicated paging systems to the use of consumer

items like mobile telephones. For example, possible options include sending emails, using instant messenger systems (e.g. IRC or MSN), playing pre-recorded, or software synthesised, messages over the telephone, or sending SMS/pager messages.

4.1.3 Publish/subscribe and broker technologies

The traditional client-server model is well suited to 1:1 relationships, where a single server transmits data between itself and a single client. This can be extended in the server software to allow multi-casting of the same data to a number of clients. The publish/subscribe model extends this further allowing for multiple data providers to communicate with multiple clients, this is achieved by passing all data through a central broker. The benefits of this are that only the location of the broker needs to be known, and that clients can then receive a combined stream containing data from a number of sources. Under a traditional client-server model, the clients would need to know the location of all the data providers, and would have to individually combine their data streams.

The publish/subscribe model is built around a central broker and a number of clients which connect to the broker. The broker acts as a go-between: an agent that matches subscribers to the publishers of information that is relevant to them. Clients can be publishers of and/or subscribers to data and can range from large enterprise-based servers to small pervasive computing devices, such as handheld PDAs, or unattended remote telemetry devices.

Publishers send units of information (called messages) to the broker on a specific topic. The topic is analogous to the subject line of an e-mail. It informs the recipient about the content of the message; subscribers register their interest in certain topics with the broker, and the broker then forwards them messages which occur on those topics. The broker meanwhile manages connections from publishers and subscribers, and deals with authentication and access control lists; controlling who is allowed to publish and subscribe to which topics.[38]

It is important to remember that publishers and subscribers are usually unknown to each other, with the broker acting as matchmaker. This decoupling helps to make the system future-proof because it generalises the use of data, which is then no longer tied to specific applications. At any time, new applications can be deployed which are able to use any combination of topics. It is this potential that makes publish/subscribe such a powerful concept.

In this work, a broker solution based around an IBM Message Queueing (MQ) Broker was used. The MQ Telemetry Transport (MQTT) protocol[39] is used for communications with and between MQ Brokers.

4.1.4 The hierarchical nature of broker topics

The MQTT specification[40] describes topics which are arranged hierarchically. It uses slashes (/) to delimit the levels in a similar fashion to a web URL. The hierarchy, which defines an information space, must be carefully designed to help ensure that data are available in a sensible, logical, structure.

In the example of environmental monitoring, we may choose to define a topic space of

```
/Project/Room/SensorType/SensorID
```

where project is an overall delimiter to separate, for example, environment monitoring from remote control, and then room and sensor type have the obvious meanings. SensorID is included to allow for more that one sensor of the same type in a room. In this scheme we may arrive at the following set of sensors when considering a simple case with a lab adjacent to an office (figure 4.1).

```
/environment/lab/temperature/1
/environment/lab/light/general
/environment/lab/light/sample
/environment/lab/door/office
/environment/lab/door/corridor
/environment/office/temperature/1
/environment/office/temperature/2
/environment/office/light/1
/environment/office/door/lab
/environment/office/door/corridor
```

The MQTT specification also defines two wildcard operator that can be used to select multiple topics from a broker. These are . to indicate a single level wildcard, and # to indicate a multi-level wildcard. By using these wild cards on this example we could select a number of potentially interesting sensor sets. An individual sensor can still be selected, by using its topic in full.

For example, we may be interested in all the sensors in the lab, in which case we would subscribe to

```
/environment/lab/#
```

Or all temperature sensors in the office

```
/environment/office/temperature/.
```

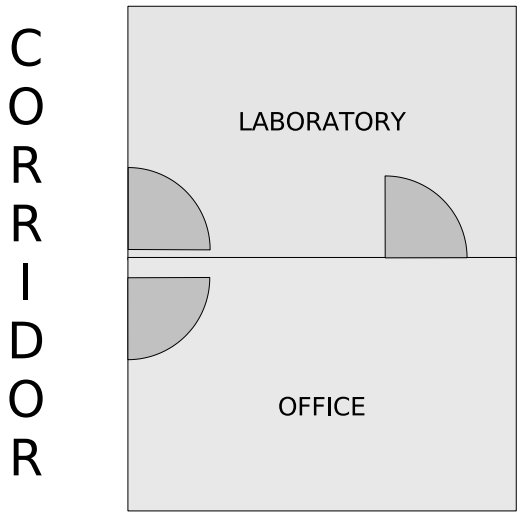


FIGURE 4.1: Room layout of the simple lab and office example.

Or indeed, all temperature sensors

```
/environment/./temperature/.
```

It should be noticed, however, that actually some of our topics are actually degenerate – there are two topics that carry the same information; it could be the case that these are separate sensors, or it could be that the data capture software replicates the information from one physical sensor onto the two topics. At first glance one might suggest that all doors leading to the lab could be selected by either of these

```
/environment/lab/door/.
/environment/./door/lab
```

However, as our topic space doesn't carry any explicit information about the corridor only the first of these would yield the complete result. Equally, given the lack of explicit corridor topics, implicit information about the corridor can be found based on topics matching

```
/environment/./door/corridor
```

When designing the topic space in this scenario, handling of degenerate topics needs to be considered; should implicit topics be included in the message space to make wildcard searching easier, or should they be specifically excluded, forcing clients to use both wild cards for cases like "door" in the example. Both methods are equally valid and the choice a matter of system policy. Care must be taken if implicit topics are only partially included; the client must account for a single search not yielding all the information but that a two part search may yield duplicates that must then be accounted for.

4.1.5 Security for broker applications

By design MQTT has a very minimalist approach to security; this allows application developers to use a security protocol appropriate to their field. MQTT is capable of carrying virtually any data, so an obvious first step to securing data in transit in a broker system would be encryption, possibly using a preshared secret, or Public Key Infrastructure (PKI) implementation to additionally provide signing for authentication and non-repudiation. Challenge/response security could be incorporated at the application level, by sending the challenge/response flows as MQTT messages over broker; this method is used internally by the broker for its administration channel.

Security can also be added at the network layer; by using strict firewall rules connectivity can be limited to only authorised machines. MQTT is a protocol based on of TCP/IP, hence standard Virtual Private Network (VPN) products can be used to secure the connection and the data flowing inside it. MQTT is often used with (Secure SHell) SSH to provide a lightweight VPN, but can also be used with more sophisticated VPN products, for example a Point-to-Point Tunnelling Protocol (PPTP) service[41].

4.2 Implementation

4.2.1 Implementation Overview

The laboratory monitoring solution used for this work was based around a publish/-subscribe network, using the MQTT protocol with an IBM Microbroker[42] provided as part of a technology collaboration with IBM's Hursley Laboratories. Data were fed into the broker system by a number of data capture devices around the School of Chemistry at the University of Southampton (referred to as "the school"), and were distributed to a variety of clients and autonomous agents. The broker hosted messages for the location awareness system, device remote control, and experiment monitoring within this project.

When the initial implementation was performed in the SHG laboratory it was decided to monitor temperature, presence of people in the laboratory (and their movement in and out) and the state of the laboratory lights (on/off). A sensor connection to the laboratory laser interlock was also made, as this provided rapid detection of unauthorised access to the laboratory (which would result in the interlock changing state, and the experiment being stopped). Values from these sensors were captured using a computer based data acquisition system, which scanned for changes in these values. When a change was detected, a message was published via the Microbroker. These messages were formatted in a XML style notation, as used by other projects[43, 44]. It was decided to include no authentication or encryption in the messaging scheme as the broker is only visible

to systems within the University of Southampton data network, and the additional programming overhead would have limited development.

4.2.2 Electronics

Temperature was measured using semiconductor temperature sensors (National Semiconductor LM35) which provided a calibrated, linear, temperature to voltage output suitable for environmental monitoring (although higher precision components may be required for sample control and experimental measurements). This signal from the LM35 is captured using a data acquisition card (National Instruments LAB PC+, LabJack UE9 or LabJack U12) connected to a personal computer (PC). The state of room lighting (on/off) was monitored by a photo-diode placed alongside one of the light fittings. The signal from this is monitored by the data acquisition card, and passed through a threshold filter to generate a software toggle that mimicked the light switch. This option was chosen in preference to a direct connection to the laboratory wiring on grounds of safety. It has the downside of monitoring only one light-fitting. A preferable solution would be to use a higher sensitivity photo-sensor placed more centrally in the laboratory, for example alongside the sample or a light sensitive piece of apparatus. In the later implementations, a light dependent resistor (LDR) was configured to form one leg of a potential divider, so providing a measurable voltage dependant on ambient light level.

Monitoring of personnel movements was performed using off the shelf security alarm components. A Passive Infra-Red (PIR) sensor in the laboratory detects people moving; it was found that due to the layout of the laboratory, and the experimental necessity for operators to remain still, the PIR ceases to detect people and stopped triggering. This was predicted during the planning stage and the ability to detect people passing through the laboratory doors was included. Door state is monitored using magnetically triggered reed switches. Additional PIR's were later added to improve coverage within the laboratory.

All these toggle sensors connect back to a binary logic input on the data acquisition card, using the card's internal 5 V supply with a bias resistor to generate the high state.

4.2.3 MQTT messages in the SmartLab

The MQTT specification[40] defines a generic messaging system, providing transport controls and a container to hold application specific messages. Using the MQTT libraries provided by IBM[45, 46] removed the need to encode the underlying MQTT messages; instead it is only necessary to produce the text of the message, and pass this to a library function. An XML messaging format, following the fragment shown in figure 4.2, was chosen. This also provides the simple messages, with low computational overhead,

needed for the smartLab approach, and was shared with other *e*Science projects at the University of Southampton [43, 44]

```
<msg>
  <data>Wiring_Box</data>
  <value>20.996094</value>
  <time>1095690315</time>
</msg>
```

FIGURE 4.2: An example of the XML messages that carry sensor information in the SmartLab system.

4.2.4 Transformation agents

A number of small "agents" are used within the SmartLab messaging system; they perform translation tasks to make the messages accessible from other systems. These can either be transferring messages away from the MQTT based system, or may be translating message from one format to another, but then encapsulating the result as an MQTT message, and publishing them back to the broker (or another broker). The agents used in the SmartLab are shown in Figure 4.3, and are described below.

Mimic Agent

The sensor system is designed to minimise network traffic by only sending messages when a change of measured value occurs. Some clients require a time dependent message stream (e.g. one message per second) and to achieve this, the mimic agent was written; it subscribes to the measurement topics from the laboratory then publishes messages once a second on mimic topics.

Bridge Agent

The Southampton broker operates within the University of Southampton data network, and access to it is restricted by the University's data access policy (as enforced by the campus firewall)[47]. A client outside the University firewall is blocked from accessing the broker directly, hence for a remote client to receive data from the laboratory it needs to be published on a publicly visible broker. The simplest solution to this would be to make the Southampton broker publicly visible (by liaising with the campus firewall team). However, this would make all internal data streams publicly visible, and could potentially allow other users on the Internet to inject data into the streams.

The solution to these problems is to bridge the data from the Southampton broker onto a publicly visible broker, in this case an enterprise broker run by IBM. The bridge agent, which is a component of the Microbroker, allows us to send only the data which we want

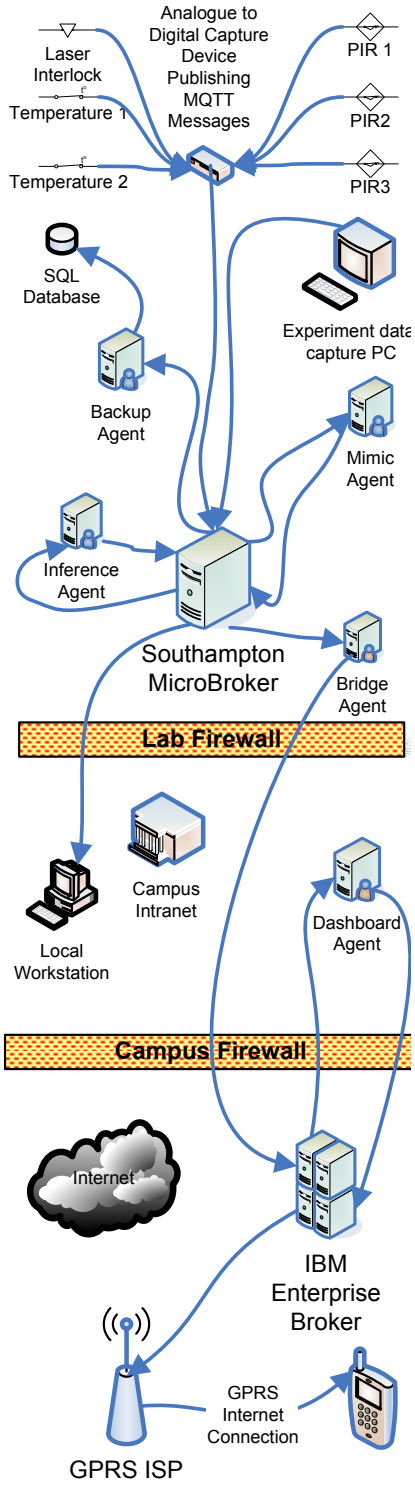


FIGURE 4.3: Data-flows within the laboratory sensor and experiment messaging system.

to make publicly visible, and is a one-way connection so external clients can't inject messages into our private data streams.

Backup Agent

Data on the broker are inherently transient in nature (or at best retaining the last message). One aim of the laboratory monitoring work is to be able to review old data so that laboratory conditions can be considered when analysing data (and possibly justifying poor data), therefore a method for retrieving old data is needed. The backup agent performs the "store" function of this recall, by subscribing to all the laboratory data streams, and writing the data from the messages into a relational database.

Recall Agent

This is the counterpart to the backup agent. Written as part of an undergraduate project[48], it listens for "recall requests" from client, retrieves the requested data from the database, and publishes it as messages via the broker.

Dashboard Agent

Some clients require messages in a specific format, which may differ from that used within the monitoring system. The IBM dashboard client (which runs on mobile telephones) is an example of this, requiring messages with a user-readable description of the conditions rather than the raw measurements. Generating these messages requires a simple transform to map messages from the laboratory topics onto dashboard topics, and to extract the data from the laboratory messages generating dashboard display texts. The agent was based on existing code provided by Andy Standford-Clark at IBM Hursley.

Inference Agent

By combining the existing data streams, and performing some action on the data, new "inferred" streams can be generated. A simple example of this is the "SHG laboratory occupied" topic. To generate this, an agent listens for messages from the three PIR sensors in the SHG laboratory, performs an OR operation on the result and adds a 30 second release delay. This provides a good approximation of laboratory occupation, as the raw PIR data is noisy (frequently toggling as people move round the laboratory), and it is rare to spend more that 30 seconds in the laboratory without triggering one of the PIR's. An extension would be to include the door data, inferring that if a PIR is triggered then someone must be in the laboratory, and that they continue to be in the laboratory until one of the doors open. At which point the un-occupied flag could be set, if someone stays in the laboratory (or enters) then their movement would soon trigger a PIR and the occupied status would be set once more.

4.3 Discussion

Following the installation of the laboratory monitoring systems, they have proved to be a useful asset although not in the way expected. From observation of the data, we have spotted instances of equipment failure, and unauthorised access, and have been able to deduce new inferences for identifying events. As yet there has been no direct correlation between faulty data and the laboratory environment data.

Whilst presenting this work at a conference, a live demonstration of the data recall was given. During this presentation it was noticed that the temperature in the lab was fluctuating far more that it would usually be expected (Figure 4.3). After the presentation an email was sent back to the researcher known to be in the lab at the time, asking what was going on. Their response was that the Air-Conditioner kept shutting down (hence giving the large rises in temperature). A service call was made, and the problem fixed.

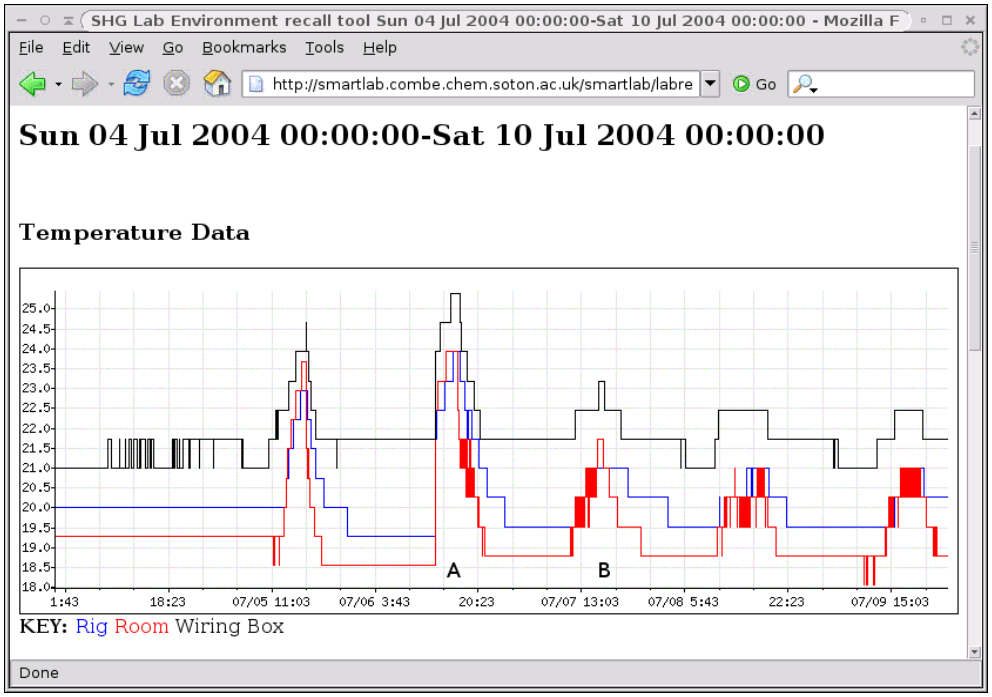


FIGURE 4.4: Environment data recall showing the air-conditioner fault. Point A is an at fault day, and Point B a normal day.

Following the implementation of the mobile phone client, an unauthorised access to the laboratory occurred and an alert was sent to the phone; resulting in that days work being retrieved with only minimal loss of time. In this case, the phone had been setup to give an alert status when the laser interlock disabled (Figure 4.5); this indicates that either it is the end of the day and the lab is being shutdown, or that someone has just entered the lab without entering the over-ride code, and hence the laser has been shutdown as a precaution. In this case the culprit is unknown, but the laser was quickly restarted,

and its warmup completed in an appropriate fashion. Had the alert not been received, then it is likely that the shutdown laser wouldn't have been discovered until the user was about to start experiments, by which time the laser would have cooled, and need to be started from scratch – losing all the experimental time.

FIGURE 4.5: The mobile phone client, in overview (left) and alert (right) modes.

Chapter 5

Location Awareness

5.1 Background

When providing the user with feedback, there may be a number of technologies and data paths available; each providing a different level of detail and possibility for interaction. Some of these may be location dependent, for example a physical dial on an apparatus, or a fixed computer screen. If the system was able to know the (approximate) location of the scientist, then it can provide feedback on the most appropriate display available.

For a number of years researchers in the semantic computing field have recognised the value of location awareness, and a number of systems for tracking individuals have evolved[23]. These are typically based around the user carrying a remote readable, unique identity tag that the system can identify and link to that individual, for example, an ultrasonic tag, a radio frequency identification (RFID) system, or tapping into some existing movement control system, such as electronic door locks (where a key or pass-code would be replaced by a unique ID and a database lookup). When outdoors there also exists the possibility to use GPS, mobile telephone cell identity or other dedicated radio triangulation technologies. Modern video processing allows for the recognition of individuals or objects in a passive, unobtrusive fashion. While automatic facial recognition may be far from reliable, the use of gait detection is increasing in popularity[49]. Where an object carries a visible unique identity tag, this can readily be captured and decoded to track that object; current examples of this include automated toll and average speed detection cameras employed on the road network in which the system is recognising the vehicle number plate in order to track its movement, for example in the M42 Active Traffic Management System[50] and the London Congestion Charging scheme[51].

These existing systems all carry a shared burden, one of cost through the need to use dedicated, bespoke hardware to perform the identity check. A brief review of the technologies carried by PhD students, Post Doctoral researchers and Academic Staff indicated that adding another dongle or tag would be unwelcome, but that they were already carrying a number of suitable tags that could be used. In the case of the University of Southampton, all staff and students are required to carry a university identity card that includes an RFID chip. This is used to provide security authentication, and indicate an individuals right to use certain facilities. However the technology used is, by design, only readable over short distances; cards need to be placed within a few centimetres of the reader. It was also noted however that most researchers now carry a mobile telephone, that the majority of modern handsets are bluetooth enabled and that a surprising number are fully discoverable by default. The bluetooth units in modern mobile telephones can typically communicate over a range of a few tens of meters in free space, which is ideal for getting an approximate room location of the device (and by implication its owner). The presence of a device set to be discoverable can be detected without the need for a formal 'handshake', or authentication - ideal for determining the device's location. There exist on the consumer market cheap PC based bluetooth devices that can perform this detection, and that come in a selection of sensitivities, further allowing for detection areas to be tuned.

Since starting this work, a number of other researchers have used the bluetooth device concept to provide location aware data, for example the NAMA project[52] uses bluetooth beacons to provide a user with location based reminders (in their example to remind the user of purchases they intend to make as they visit the appropriate retailers), and the LAMA project[53] which provides location specific information direct to mobile phones (for example platform alterations in train stations). However, both these systems are designed to allow a user to discover their location without necessarily relaying this information back to the information provider. Gonzalez-Castano *et al* propose a server oriented system in their 2005 paper[54] where a number of nodes detect the presence of their clients and relay this information back to a master server, which then pushes information out to the clients relevant to their location (using a parallel WIFI network), in their example this is information sheets in a museum situation. The Asterisk[55] open telephony platform has a number of modules which can use Bluetooth as a presence detector, for example to set up a call divert from an office phone to mobile when a user leaves the office[56]. The LOCA project[57] is also of interest, it discusses implementation a network of passive Bluetooth scanners that report the location of nearby devices back to a central server, the users of the detected devices are then invited to take part in social interaction experiments that highlighted the issues of device security in modern technology.

5.2 Implementation

In order to provide a cost effective and easily implemented location awareness system it was decided that the use of bluetooth devices was preferred; read distances of up to 100 metre are possible with inexpensive consumer USB dongles, which is more than adequate to cover a laboratory.

Bluetooth dongles were deployed to the Linux PC's in the rooms to be covered. A daemon was created that periodically (every 30 seconds) performs a bluetooth device discovery. It then compares the list of currently visible devices to a list of devices seen at the previous scan. Differences between the two lists are then published via the broker; a "device discovered" message (Figure 5.1) is sent for all new devices, and a "device lost" message is transmitted for devices not appearing on the currently visible list. The daemon is also able to respond to requests (sent by MQ) asking for either messages about currently know devices to be resent, or for the location of a specific device to be revealed. This behaviour is shown in Figure 5.2.

```
<msg>
  <type>Device_Found</type>
  <beaconID>rutherford</beaconID>
  <clientMAC>11:22:33:44:55:66</clientMAC>
  <clientName>Happy_Camper</clientName>
  <time>1095690315</time>
</msg>
```

FIGURE 5.1: Example of the messages that carry location information.

These messages allow for a database to be maintained, listing the times that a device was visible from a specific location and allowing devices to be found on demand (Figure 5.3). The storage action is performed by a second daemon that listens for messages from the beacons and updates the database accordingly. This behaviour is shown in Figure 5.4. Any application within the system that needs to know the location of a user can then either query the beacons directly or query the database.

5.3 Ethical Implications

When handling data about people rather than those relating to experiments additional care has to be taken to ensure that the people involved are aware of, and in control of, how their data are being used. This is equally, if not more, important in cases where these data are being captured covertly, such as with the Bluetooth tracking infrastructure described here.

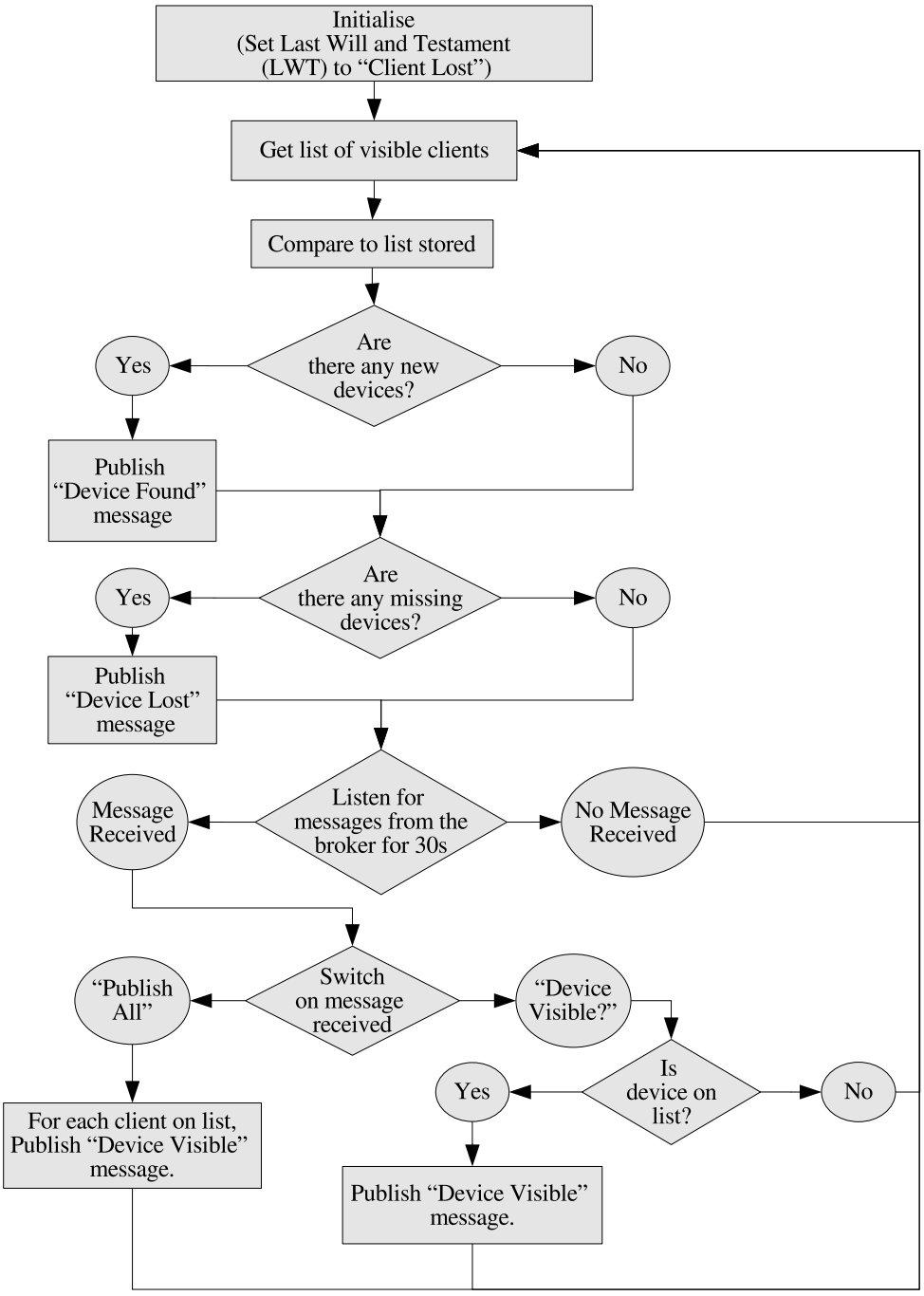


FIGURE 5.2: Data flow within the bluetooth beacon client.

MAC Address	11:22:33:44:55:66
Device Name	Happy_Camper
Time Found	1190625010
Time Lost	1190647038

FIGURE 5.3: Data stored in the location database.

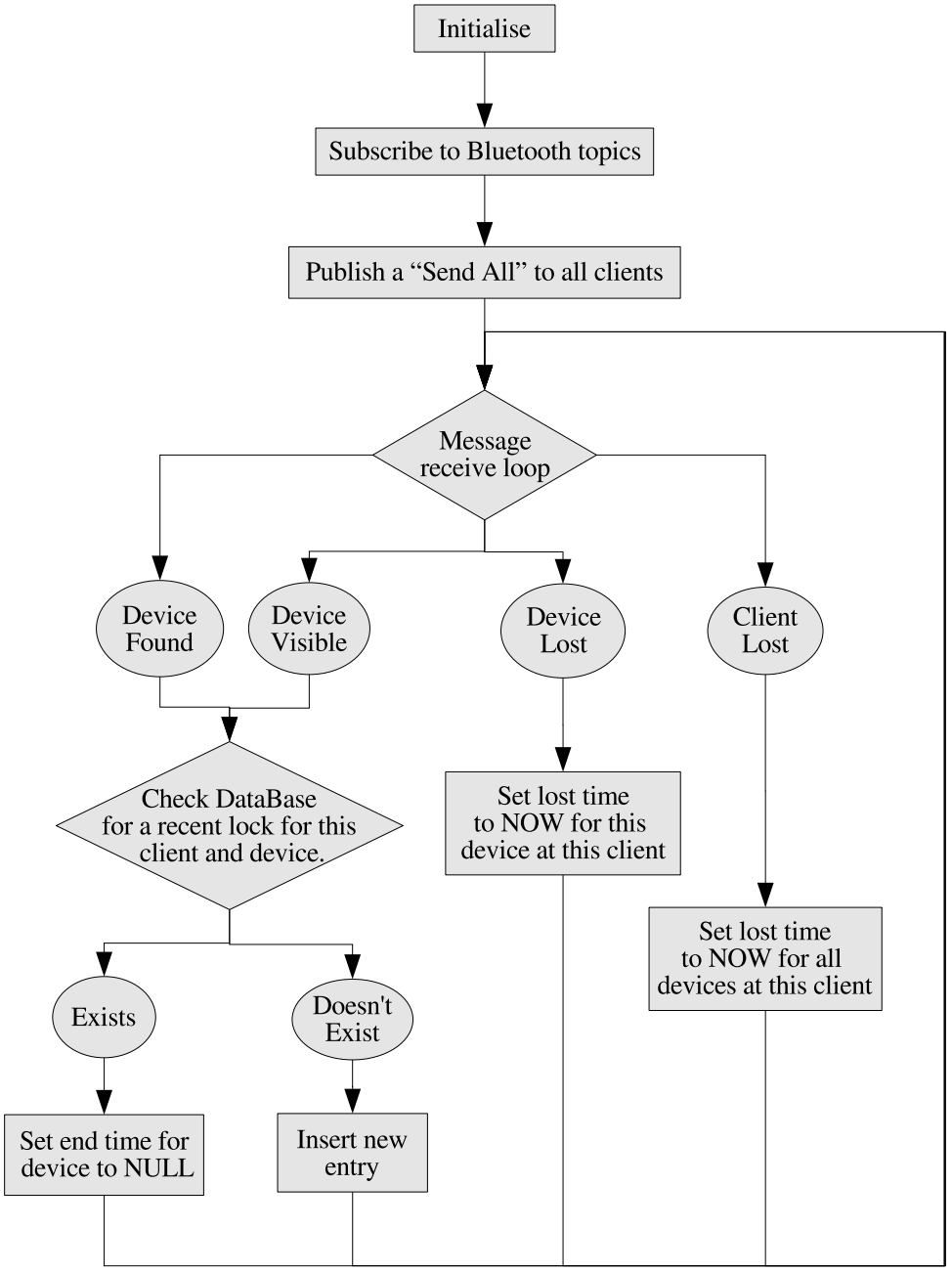


FIGURE 5.4: Data flow within the bluetooth database manager.

The system detects all Bluetooth devices in the vicinity which are responding to a Bluetooth discovery request, and logs the location of these devices in a database. This database can then be searched for known device identifiers which correspond to the individuals to be tracked. The location of such devices is then recorded against that person's profile. The system can only locate those individuals who have specifically requested this service by providing the identifiers of their devices. During development, the group of users is limited to those people involved in the system's development, and hence who have made their own judgements about the security of the system and how it

affects their privacy. Before enrolling members of the wider community to the system, an assessment of the system's data usage must be made, and statement of that assessment produced so that the users can be adequately briefed.

In addition to the managerial controls, there also exists a technological control; only those devices which are configured to be discoverable can be found, so, to avoid being tracked one has only to set their device to be non-discoverable. Hence, any user who is aware of the system is in full control of when their current location data are used. However, this does not cover the storage of past location records or the detection of unknown devices, probably belonging to people who are unaware of the system.

In the case of unknown devices, it was initially assumed that the identity retrieved from the device is random and doesn't provide an obvious link to the owner, and secondly that people realise the implication of setting a device to be discoverable. In practice both of these assumptions are flawed. Bluetooth devices are identified in the system by two characteristics, firstly the network hardware has a unique identifier, its hardware (MAC) address, and secondly the user can assign a name to help them identify the device remotely. The MAC address is a quasi-unique hexadecimal string based on a manufacturer ID and usually the order off the production line of the network chip; and as such bears no link to its owner, other than by association. The user set name, however, usually provides a much stronger link to the owner. In the best case for confidentiality, it would also be a seemingly random string, but one with some meaning to the owner and only the owner. Usually, however, they default to being the make and model number of the device (e.g. "nokia 6210"), or when set by the user tend to be quite obvious (e.g. "Andy's Phone"), occasionally the user will select something which is less meaningful to anyone else (for example a fictional character or pet's name). Whilst running this project the MAC address and friendly name of all the Bluetooth devices observed have been added to a database. A review of all the devices seen (some 234 devices) shows a spilt of 21 % devices on default name (typically identifying a device by model), 41 % using an identifiable name (e.g. "Andy's Phone") and 38 % using a name not readily identifiable.

The setting of device visibility is a choice the user makes when using their device. However, it is a choice they are forced into to get functionality – a device must be discoverable before it can be discovered and paired with an authorised host. Once this pairing has been completed, it is usually possible to disable the discoverable feature without loosing the functionality of the pairing. This is, however, another step, one that most users overlook either through ignorance or laziness, so leaving their devices discoverable. It is unreasonable, therefore, to assume that just because a device is discoverable its owner wants to be discovered.

The LOCA project[57] highlighted this ignorance of passive surveillance by leveraging the broadcast nature of Bluetooth identity and responding to discoverable nodes with

provocative messages, initially to gain a visible response (for example sending a message asking the user to wave their device in the air), and then to raise awareness of the information that the users were giving away without realising. They found that many people are unaware of how easy it is to track a person's movements with simple off-the-shelf hardware, and equally that by simply setting a device to be un-discoverable it is made far harder to track.

5.4 Current Usage

Data from the Bluetooth beacon network isn't currently used directly by any projects, but feeds into a larger location awareness database. This database is currently used by two opt-in services used within the group. The first of these uses is a printer awareness service, used by the chemtools blog[58]. The blog produces a unique identifier for each entry, that can be represented as a barcode to be stuck into a traditional paper lab book or onto a sample and uses the location database to automatically select the nearest barcode printer when a request for a barcode is received. The second use is to relate a user's position to other colleagues. In the current implementation, the user's current most recent know location (derived from a number of sources including the Bluetooth beacon network) is plotted onto a Google maps page, along with details of the time, and data source. Both of these services require the user to actively opt in, giving their permission both to be tracked and to have their Bluetooth devices associated with them.

Chapter 6

Remote Control

6.1 Background

In chapter 4 the benefits of the chemist not being present in the laboratory during data collection (both for safety and quality of results) were discussed. By providing the chemist with the ability of control aspects of the experiment remotely the time spent in the laboratory can be further reduced, allowing the experiment to run even longer without human presence. The term "lights-out operation" is frequently used to describe equipment which is controlled remotely or runs autonomously.

Remote control can be both beneficial and problematic for laboratory personnel if the equipment or process is inherently dangerous, for example if it contains moving parts, or radiation sources. Remote control allows for personnel to reduce their exposure to the danger by not being present however a remote operator must be fully aware of the state of the laboratory, so as not to put persons working in it at risk by triggering the equipment unexpectedly. The use of local mechanical and electrical interlocks can help to avoid the consequences of a remote action, and by adding extensive remote monitoring the remote operator is placed in a more knowledgeable position. In the design of such a system one also needs to take account of the possibility of malicious activities. The broker implementation used in this work is incomplete as authorisation and access control were purposely ignored. If a malicious user was to monitor the broker channels during a period of remote control activity they could derive the command set required to access the apparatus and then replay these commands at will, either in an attempt to disrupt experiments, affecting the results collected, or to cause injury to a local user. In the university research environment the motivation for such an attack is quite low, however in an industrial setting the competitive market place may cause such motivations to be much more significant. A local operator could use the same remote monitoring and control technology for local control, and if the system were aware of

their presence in the lab they could be given a higher level of access than a truly remote user.

The broker system implemented to facilitate remote monitoring (see chapter 4) can also be used as the communication channel for driving a remote control system. In the case of the SHG experiment there is both a risk of injury to the operator and risk of damaging the equipment if inappropriately timed commands are sent to the apparatus. It was decided that a simulated lab environment would be more appropriate for investigating use of the technologies. In an unpublished work, Andy Stanford-Clark and colleagues at IBM's Hursley Laboratories demonstrated the integration of LEGO MindStorms with an MQTT broker infrastructure, to monitor the MindStorms' sensors, and control the MindStorms' actuators. Their work focused on a number of small scenarios where the networking was assured, and all data movement was via the broker. In the laboratory scenario, networking is likely to be reliable but will not always be so, and there is a need to deal with data travelling by alternate streams, such as experimental data or video feeds from cameras. There is also the likelihood of data being delayed in transit, for example buffering of video streams.

6.2 Authorisation and access control

When providing remote access to a service, authorisation to use that service must be considered. It may be decided that anyone should be able to use the service (as for most information available through the Internet), in which case there is no need to implement any access control. However, as soon as a need arises to identify users or restrict access to a service, then some form of access control must be considered. There exist a range of options for identifying the user, with varying degrees of certainty they really are that user. These range from simple username logon, through certification based authentication to hardware based authentication methodologies, not forgetting a large range of potential biometric identifiers. Typically, these systems produce some form of token that can be checked against a store of trusted, or untrusted, tokens, so allowing an authorisation decision to be made; the exact nature and location of these components is what varies in the various authorisation methodologies.

In a simple username/password system one could conceive an approach where the service being accessed holds a table of authorised users and their authentication token (password). To access the service, the user would be prompted to provide their username and password, and when these matched the values held by the service that user would be granted access. This system is straightforward to comprehend and provides a low to moderate level of certainty that the user is who they claim to be - in a perfect world only the user would know both their username and password, hence a good level of certainty that the user was authentic would be achieved however username/password pairs are

easily shared, and are relatively trivial to recover from unencrypted network traffic since as with any shared token, it is transmitted over the wire each time the user is authenticated. The storage of tokens by the service also makes this username/password pair useful only for that service. Whilst a user may choose to set the same pair in multiple places, there is no automatic synchronisation of this information.

Various techniques have been applied to extend this simple scenario to overcome some of its shortfalls. The fundamental problem of users sharing of username/password pairs cannot easily be overcome, however the use of encryption removes the triviality of recovery from network traffic. Some work has been done recently to overcome the disparate password database problem, leading to systems for single sign-on; within an organisation this can be achieved by giving all the services access to a central username/password database (for example using the Lightweight Directory Access Protocol (LDAP), or Active Directory). The technology used can be extended to remote sites, however, politically there is an issue of trust when granting a remote site access to user credentials and information, as the remote site could use them to gain access to any services those users can access. This problem can be tackled in one of two ways, firstly by restricting access to only those parts of the password database the remote site needs to perform its service the scope of data misuse is minimised, or secondly by an authentication hand-off whereby the user is authenticated by a service running at their home site which then passes confirmation of this authorisation across to the remote service. The latter method ensures that the remote site never gains access to the user's credentials, and instead only receives the information needed to provide their service. Such a system is implemented by the Shibboleth project[59], which is currently being implemented by UK universities for access to shared resources.

Within the *e*Science community, there has also been some talk of Virtual Organisations (VOs); groups of people who have a common goal, but are not linked by a traditional organisation, for example an inter-institutional or interdepartmental research project as is common amongst the UK *e*Science projects. Such a VO could be authorised by a double layered version of the shibboleth model described above; a service provider may want to allow all members of a VO to access their service and would then trust that VO as an authorising agent. The VO would not maintain its own authorisation tables, but would maintain a list of its members and the authorising agent at their traditional, physical organisation, hence that VO can then pass through the authorisation procedure to the user's physical organisation.

6.3 Implementation

In order to test the concept of remote control over the broker it was decided to build a test case using LEGO MindStorms. An IntelPlay microscope was mounted into a LEGO

FIGURE 6.1: The IntelPlay microscope mounted in the LEGO frame.

framework and a system of LEGO mechanics was used to provide a sample positioning system, bringing the sample in and out of the focal plane of the microscope. This work can be broken down into four discrete sections, namely, physical construction (the LEGO bricks, figure 6.1), RCX software (the link between MQTT and the LEGO sensors and motors), microscope software (controlling the camera hardware) and the remote control front-end. These individual components are discussed in the following sections, and their relationships shown in figure 6.2.

6.3.1 LEGO construction

The LEGO construction around the microscope needs to support the microscope and sample, to provide sample selection/movement and focusing ability. The initial design involved focusing the sample by translating the microscope vertically but the weight of the microscope and quantity of available LEGO dictated that it was more appropriate to move the sample vertically below the microscope, keeping the microscope static. The sample platform was driven via a rack and pinion lift powered through reduction gearing by a MindStorms motor. A combination of gearing and driving the motor for very short time periods allowed the fine control needed to bring the sample into focus even at the highest magnification settings. For protection, the sample carriage was equipped with vertical limit switches. Both limits (top and bottom of movement) were connected to the same sensor input by being wired in parallel and which limit had been reached was inferred from the direction of travel. In a further experiment, the microscope was mounted on a LEGO tractor which could be driven over a sample mounted on the floor.

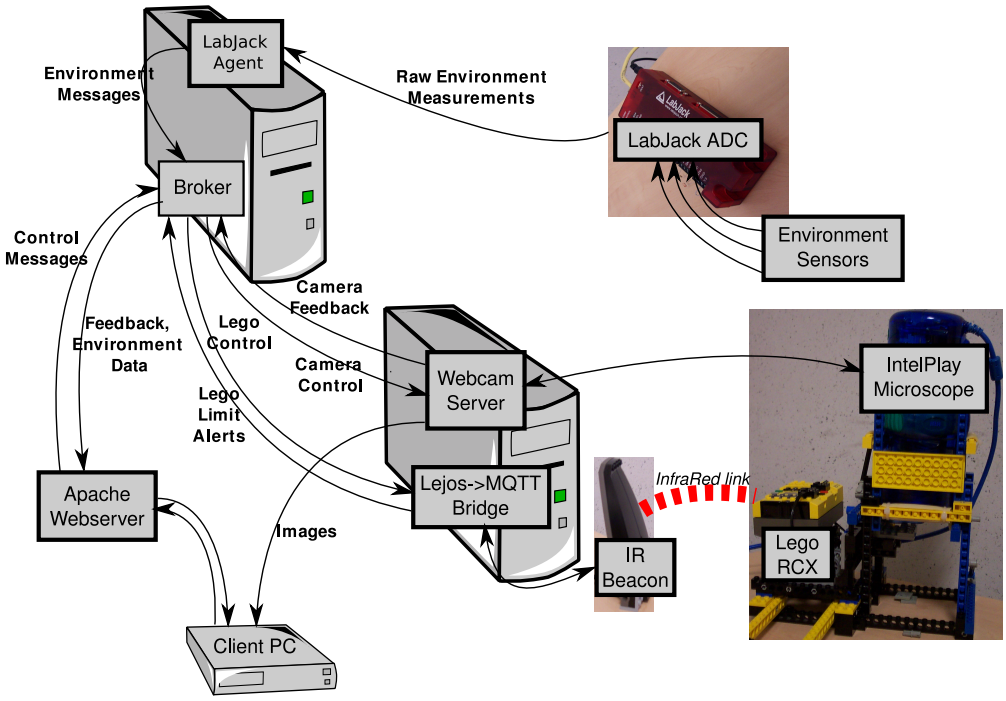


FIGURE 6.2: Block diagram to represent the software interactions within the MQTT controlled LEGO microscope demonstrator.

6.3.2 RCX Software

The LEGO RCX drive software was split into two parts. Both of these parts originated from IBM Hursley and were modified for this project. The LEGO IR tower and the IR receiver on the RCX are used to link the RCX to the MQTT network in real-time, however due to the limited processing capabilities of the RCX a full MQTT protocol stack is not used. The implementation used sends single integers to trigger commands on the RCX and a two integer pair to receive a topic and message from the RCX for publishing over MQTT.

This solution requires two software components. One runs on the RCX, and one on the host PC which bridges between the messages on IR link and MQTT messages over a TCP/IP network. To allow more extensive programming of the RCX, the standard LEGO firmware is removed and the "leJOS" system[60] installed instead. LeJOS provides a basic Java-based programming environment on the RCX, as well as PC-side libraries to enable communication with the RCX over the IR link. This makes programming interactions between MQTT and the RCX easier, as the standard MQTT Java libraries can be used on the PC side of the link. The PC-based bridge subscribes to a topic from the broker and, on receiving messages, retransmits the payload over the IR link (after filtering the message payload to contain only integer values). The software on the RCX receives messages over the IR link from the PC and if the integer received

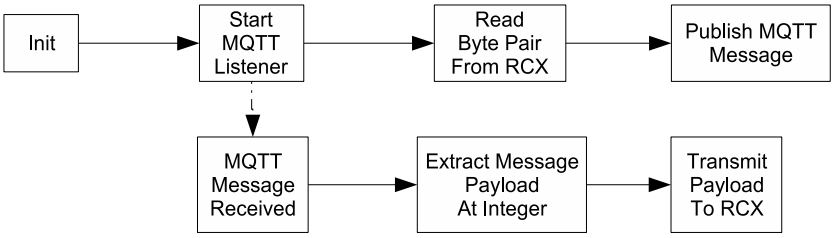


FIGURE 6.3: Software flow within the RCX PC bridge agent.

corresponds to one of the expected control codes, the appropriate action is taken, otherwise the message is ignored. On receiving a sensor event (the leJOS code allows event triggers to be tied to changes in sensor state) the RCX may perform simple processing on the data, triggering actions as required, and then transmitting a message over the IR link to the PC. Figure 6.3 shows program flow within the PC agent.

In the case of the LEGO Microscope, there is presently one motor attached to motor control A which drives the focus position, and two limit switches (one for each end of travel) connected in parallel to sensor input 1. The RCX accepts commands to stop the motor, and to run it forwards and backwards. It also accepts a command to "bump" the motor - this drives the motor for 50 ms in the direction requested, then stops it. When the limit switches are triggered, the code firstly stops the motors, then backs the motor off in the opposite direction for 200 ms (to release the limit condition), it then publishes a message over the IR link stating which limit switch has been triggered.

Figure 6.4 shows this program flow schematically.

6.3.3 Microscope software

The IntelPlay microscope is viewed from the computer as if it were a web-cam, using a CPIA chip-set[61]. The author of the Linux driver made control of the camera's internal processing easy through the use of the Linux /proc interface[62]. Through this interface we are able to read and control sample lighting, and image brightness, colour saturation and hue.

In the current system, images are transmitted from the microscope using a simple web-cam package[63]. This presents images over an HTTP connection as either JPEG static frames or as a multi-part JPEG. Control of the cameras internal processing and sample illuminator is provided by the camera driver. A java MQTT client subscribes to a microscope control topic on the broker, and on receiving control messages writes the appropriate commands into the camera's control interface.

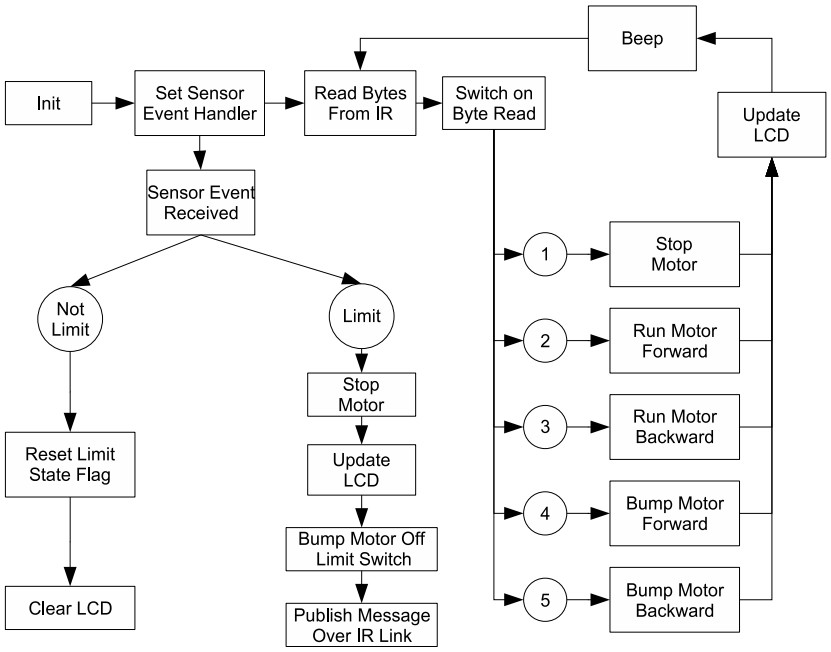


FIGURE 6.4: Procedure flow within the RCX Code.

6.3.4 Frontend control

By using MQTT to pass messages within a remote control system front-end, interfaces can be developed using a range of programming languages on a variety of platforms. For this demonstrator, the remote control interface was presented as a web-page. Within this page, the user is presented with apparatus controls using HTML form elements alongside images from the microscopes camera.

MQTT control messages are generated using a Perl CGI script running on the webserver. When the end user triggers one of the control buttons on a webform, a CGI script is called which connects to the broker and publishes messages for the RCX or microscope software.

6.4 Discussion

The decision to use MQTT as part of the control infrastructure for a remote controlled experiment was reached based on experience of using MQTT for remote monitoring. The broker architecture had already proved to be reliable and swift when delivering messages. Familiarity with the product helped considerably when implementing the remote control infrastructure, greatly reducing development time when compared to implementing a full client-server based control system.

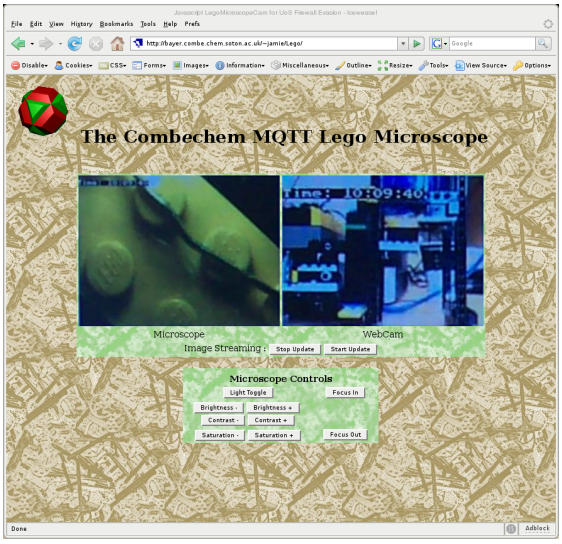


FIGURE 6.5: The end user interface to the microscope.
Showing the camera controls (left) and LEGO controls (right).

The broker allows for multiple persons to control the apparatus simultaneously, by seamlessly combining the messages into a single control stream to be passed to the hardware. This was not without problems in the implementation described here as there was no authorisation or priority system, hence, when there were multiple simultaneous users present it was common for a "fight" to occur over which action the apparatus was to perform. It was noted that the web-cam streaming system used introduced a reasonably large delay (in the order of two or three seconds) and, as this was the only user feedback, the feeling of distance involved was exacerbated. Users often commented that getting the final fine focus correct was somewhat hampered thus and it was common for them, in their impatience, to send multiple move commands and go past the point of best focus. The delay in feedback also led to control difficulties when there were multiple users. If there had been instant feedback to the users that another action had been received, then they may be more patient in waiting to see the nature of what the image was changing to rather than going ahead with their own plan of movements. A software mimic of the expected behaviour of the apparatus would provide some level of realtime feedback, and by generating this based on a subscribing to the control topic on the broker the simulation would show the result of all messages sent.

It was found that the IR link between the PC and the RCX was somewhat unreliable, especially in bright light and as the distance between the IR tower and the RCX increased. It was quite common for a message not to be delivered immediately. However, because of the persistence and quality of service within the MQ protocol, these messages were always delivered eventually; this could have led to potential problems had this been a real experiment. If an action triggered late or unexpectedly the risk of injury to local personnel could have been quite significant. MQ also provides other quality of service

options, and in the case of remote control it may sometimes be more appropriate to set messages to attempt delivery once only, and in the case of failure to discard the message optionally notifying the sender of the failure.

Chapter 7

Second Harmonic Generation
The Chemistry Experiment

Experimental second harmonic generation, the monitoring of a material's ability to produce radiation at double the frequency of light incident on it, was made possible in the early 1960's by the invention of the LASER, and was first reported by Franken *et al.* in 1961[64]. In their experiment, a Ruby Laser (fundamental 694 nm) was fired through a sample of quartz, and the beam emerging from the the quartz was shown to contain a small amount of light at 347 nm. This initial demonstration of an optical frequency doubler led the way for the development of other doubling crystals, such as those used in modern laser systems. Later work by Freund in 1967[65] showed that second harmonics could also be generated at liquid surfaces, providing the basis of the technique used in this work.

7.1 Background theory

This section assumes basic familiarity with the electromagnetic nature of light (Glossary Item 15 on page 122), polarisation (Glossary Item 17 on page 123) and principle of refractive index (Glossary Item 18 on page 123).

7.1.1 Basic theory of SHG

When a light wave is incident on a dielectric material, the electric field of that wave causes a distortion in the charge distribution within that material. This distortion causes an induced dipole moment, which re-radiates at the same frequency (as shown by the law of conservation of energy). The sum of the dipole moments per unit volume

is called the electric polarisation, **P**, and is proportional to the electric field, vector **E**, hence:

$$\mathbf{P} \propto \mathbf{E} \tag{7.1}$$

Assuming the one dimensional case, **P** and **E** become scalar. The proportionality constant in this case is the product of $\chi^{(1)}$, the electric susceptibility of the medium, and ϵ_0, the permittivity of free space[66], giving

$$P = \epsilon_0 \chi^{(1)} E \tag{7.2}$$

At higher light intensities (and hence higher electric field strength) non-linear effects can be observed. These can be described by a power series expansion:

$$P = \epsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + ...) \tag{7.3}$$

where $\chi^{(2)}$ and $\chi^{(3)}$ are the second and third order electric susceptibilities of the medium. These give rise to the generation of double and triple frequency radiation. By describing the electric field associated with a light wave as a cosine function dependent on frequency, it can be written:

$$E = E_0 cos(\omega t) \tag{7.4}$$

where ω is the frequency of the light, t is time and E_0 describes the electric field at t=0.

Hence, when the light wave is incident on a dielectric medium, combining equations 7.3 and 7.4 gives the resulting polarisation:

$$P = \epsilon_0(\chi^{(1)}E_0 cos(\omega t) + \chi^{(2)}E_0^2 cos^2(\omega t) + \chi^{(3)}E_0^3 cos^3(\omega t) + ...) \tag{7.5}$$

Then, considering just the second order polarisation term:

$$P^{(2)} = \epsilon_0 \chi^{(2)} E_0^2 cos^2(\omega t) \tag{7.6}$$

and substituting in the trigonometric relationship between $cos^n(x)$ and $cos(nx)$ gives:

$$P^{(2)} = \frac{1}{2}\epsilon_0 \chi^{(2)} E_0^2 [cos(2\omega t) + 1] \tag{7.7}$$

This shows that there is a component of the second order polarisation which oscillates at twice the frequency of the incident light, implying an emission at this frequency.

7.1.2 Nature of second harmonic generated emissions

We can also see from equation 7.6 that the intensity of the second harmonic emission is dependent on the second order electric susceptibility of the material and the square of the intensity of the incident radiation. It must be remembered that second order emissions only occur at large incident intensities as the second order susceptibility is small compared to the first order susceptibility, but that these intensities are within the range of modest modern laboratory laser systems.

Second harmonic emissions only differ from their parent radiations in their polarisation and frequencies, everything else is identical. The use of coherent light sources such as laser beams results in the emitted radiation having identical coherent properties, hence the second harmonic radiation can be distinguished from other two-photon processes, such as fluorescence. This allows for the use of second harmonic active crystals as frequency doublers in laser systems. Second harmonic generation is also an instantaneous process, unlike many other multi-photon phenomena[67].

7.1.3 Origins of the liquid surface specificity of SHG

Equation 7.1 shows that an electric field, E_ω, will induce a polarisation, P_ω, in a dielectric medium. It follows that for a material with a centre of inversion, an electric field in the opposite direction, $-E_\omega$, will induce a polarisation in the opposite direction $-P_\omega$. By considering the second harmonic generation case, equation 7.7 shows that an electric field, E_ω, will induce a second harmonic polarisation, $P_{2\omega}$, hence when the material has a centre of inversion the opposite electric field$-E_\omega$ would be expected to give polarisation in the opposite direction, $-P_{2\omega}$. However, equation 7.7 shows that the only time this holds true is when $\chi^{(2)} = 0$, which indicates no second harmonic is produced.

Bulk liquids and gases are, by definition, isotropic, implying that they have a centre of inversion. Hence, for bulk liquid and gas no SHG occurs. However, the interface between two isotropic regions is anisotropic so second harmonic generation becomes viable. This also allows for the probing of liquid-liquid interfaces, as the bulk liquids don't affect the signal (assuming that the liquid is transparent at the fundamental and doubled frequencies).

7.1.4 Intensity of the second harmonic signal

In their 1988 paper, Mizrahi and Sipe[68] derived the following equation linking the SHG signal to experimental parameters for the reflection configuration, as used for this work:

$$I_{2\omega} = \frac{32\pi^3\omega^2}{c^3 A} sec^2\theta_{in} \left| \hat{\mathbf{e}}_{\mathbf{2}\omega} \cdot \chi^{(\mathbf{2})} : \hat{\mathbf{e}}_\omega . \hat{\mathbf{e}}_\omega \right|^2 I_\omega^2 \tag{7.8}$$

where the subscripts ω and 2ω indicate parameters calculated for the fundamental and second harmonic radiation respectively. I is the intensity of the radiation, θ_{in} is the angle between the incident beam and the surface normal, c is the speed of light, and $\hat{\mathbf{e}}_\omega$ is the product of the incident polarisation vectors and their corresponding Fresnel coefficients. χ^2 is the second order susceptibility, as before.

The vector $\hat{\mathbf{e}}_\omega$can be broken down into two components:

$$\hat{\mathbf{e}}_\omega = \bar{\mathbf{e}}_\omega . \mathbf{e}_\omega \tag{7.9}$$

where $\mathbf{e}_\omega$ is the Fresnel coefficient that relates the incident electric field to that in the interface, and $\bar{\mathbf{e}}_\omega$ the unit vector describing the polarisation of the incident radiation field.

It is convenient to describe the vector $\hat{\mathbf{e}}_\omega$ in reference to the Cartesian coordinate system, with components E_a where a=X,Y,Z. However, light is usually described relative to the two mutually orthogonal polarisations, s and p, which are perpendicular and parallel to the plane of incidence of light. To link values in the two systems, a laboratory frame of reference needs to be defined. It is typically defined with the Cartesian Z axis being the surface normal. The XZ plane is then aligned such that it coincides with the plane of incidence. Hence, the s polarised beam oscillates parallel to the Y axis, and the the p polarised beam in the XZ plane.

Therefore, we can express the vector $\bar{\mathbf{e}}_\omega$ in equation 7.9 above in terms of s and p polarisations, giving:

$$\bar{\mathbf{e}}_\omega = \begin{pmatrix} s_\omega \\ p_\omega \end{pmatrix} = \begin{pmatrix} \sin\gamma \\ \cos\gamma \end{pmatrix} \tag{7.10}$$

where γ is the polarisation of the incident light in the sp plane.

By combining this with $\mathbf{e}_\omega$ in component form, equation 7.9 can be expressed:

$$\hat{\mathbf{e}}_\omega = \begin{pmatrix} E_{X\omega} \\ E_{Y\omega} \\ E_{Z\omega} \end{pmatrix} = \begin{pmatrix} \cos\gamma e_{X\omega} \\ -\sin\gamma e_{Y\omega} \\ \cos\gamma e_{Z\omega} \end{pmatrix} \tag{7.11}$$

The term $\chi^{(2)} : \hat{\mathbf{e}}_\omega . \hat{\mathbf{e}}_\omega$ in equation 7.8 contains the interaction of the susceptibility tensor with the incident light, and is referred to as the nonlinear polarisation tensor. It has been shown[66, 67] that by expanding $\chi^{(2)}$, a third rank tensor, the following holds:

$$\mathbf{P^{NL}} = \begin{bmatrix} 0 & 0 & 0 & \chi_{XYZ} & \chi_{XXZ} & 0 \\ 0 & 0 & 0 & \chi_{XXZ} & -\chi_{XYZ} & 0 \\ \chi_{ZXX} & \chi_{ZXX} & \chi_{ZZZ} & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} E^2_{X\omega} \\ E^2_{Y\omega} \\ E^2_{Z\omega} \\ 2E_{Y\omega}E_{Z\omega} \\ 2E_{X\omega}E_{Z\omega} \\ 2E_{X\omega}E_{Y\omega} \end{bmatrix} \tag{7.12}$$

7.1.5 The effect of polarisation on second harmonic signal

By combining equations 7.8, 7.11 and 7.12, the following have been derived that relate to s out and p out polarisation of SHG[8, 66]:

$$I_p^{(2\omega)} \propto \left| A\cos^2\gamma + B\sin^2\gamma \right|^2 I_\omega^2 \tag{7.13}$$

$$I_s^{(2\omega)} \propto \left| C\sin 2\gamma \right|^2 I_\omega^2 \tag{7.14}$$

$$A = a_2\chi_{XXZ} + a_3\chi_{ZXX} + a_4\chi_{ZZZ} \tag{7.15}$$

$$B = a_5\chi_{ZXX} \tag{7.16}$$

$$C = a_1\chi_{XXZ} \tag{7.17}$$

where the coefficients a_1 through a_5 are the Fresnel factors, which are functions of the refractive indexes, the angles on incidence, reflection and transmittance, and the electric field components at the surface. Their derivation is discussed in the literature, for example by Higgins, Corn *et Al* [69], or Brevet [70]. This derivation has been previously used by the group[66, 71], and the calculation of the coefficients wrapped into a FORTRAN program that, when provided with the experimental parameters, produces the Fresnel factors.

From the equations for the s and p output polarisation of SHG (equations 7.13 and 7.14), the equations for output polarisation of 45^o and -45^ocan be derived :

$$I_{(+45)}^{(2\omega)} \propto \left| \frac{1}{\sqrt{2}} (A\cos^2\gamma + B\sin^2\gamma + C\sin 2\gamma) \right|^2 I^2 \tag{7.18}$$

$$I_{(-45)}^{(2\omega)} \propto \left| \frac{1}{\sqrt{2}} (A\cos^2\gamma + B\sin^2\gamma - C\sin 2\gamma) \right|^2 I^2 \tag{7.19}$$

By collecting data for s, p, and 45^o output polarisations and fitting them to equations 7.13, 7.14 and 7.18, the coefficients A, B and C can be found for a particular sample. Once these are known, the susceptibility tensor can be re-constructed using the following re-arrangements of equations 7.15, 7.16 and 7.17 :

$$\chi_{XXZ} = \frac{C}{a_1} \tag{7.20}$$

$$\chi_{ZXX} = \frac{B}{a_5} \tag{7.21}$$

$$\chi_{ZZZ} = \frac{A - \left(\frac{a_2 C}{a_1}\right) - \left(\frac{a_3 B}{a_5}\right)}{a_4} \tag{7.22}$$

These values can then be related to the energy minimised molecular structures, and information about molecular orientation at the interface can be resolved[66].

7.2 SHG optics

The SHG apparatus used was as per figure 7.1.

The laser source was a Continuum MiniLite, a Nd:YAG laser with internal doubling crystal, producing 5 ns pulses at 532 nm with a repetition rate of 20 Hz.

A reference measurement of raw laser power was taken by splitting away 8 % of the beam using a glass slide, placed at 45^o to the beam path, and directing this into a photo-diode with appropriate neutral density filters fitted to reduce the beam to withing the diode's range. The remainder of the beam (approx 10 mJ per pulse) was then trained onto the SHG apparatus at 60^o to the surface normal. The beam first passed through a half wave plate (A) and splitting polariser, to provide fine power control and reduce power down to experimental levels (between 0 and 5 mJ per pulse). The beam then passed through a second half-wave plate (B), giving control over the polarisation of the light incident on the sample, a focusing lens, and finally a yellow Schott UV filter, to remove any 266 nm light generated in the optics, before hitting the sample.

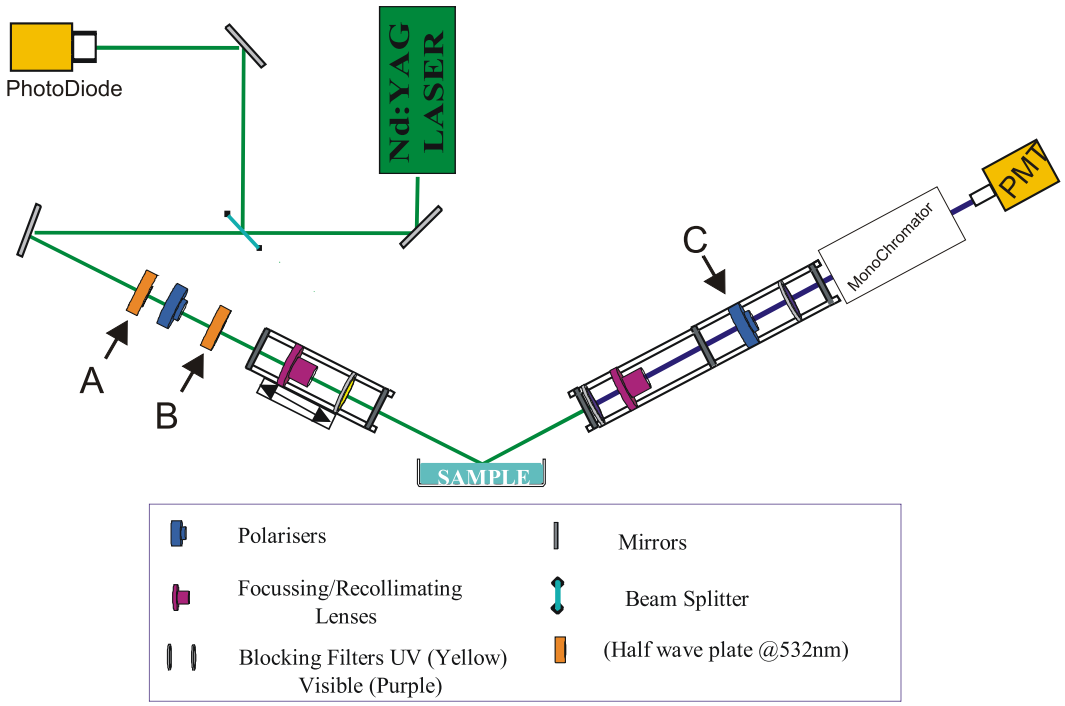


FIGURE 7.1: Optical arrangement used in this work.
Based on Diagram by Lefteris Danos[66]

The beam leaving the sample passed firstly through a UG-5 Filter to remove all 532 nm light, and hence prevent generation of 266 nm light in the optics. It then passed through a rochon polariser (C), to allow selection of the polarisation of light being measured. The beam passed through a final focusing lens before entering a monochromator (PTI International, grating blazed at 300 nm with 1200 lines per mm, and a focal length of 200mm, resulting in a first order dispersion of 4.167 nm $mm^{-}1$), and finally being incident on the detection Photo-Multiplier Tube (Thorn-EMI R166, with Thorn-EMI A2 preamp).

The sample is held in a PTFE or Glass petri dish, placed within a Perspex box (with suitable cutouts for the optical equipment) to minimise disturbance from air flow in the lab.

7.2.1 Optics zero alignment

Computer software or hardware faults have the potential to cause the home position of the automated optics to be lost. This home position corresponds to the optic having no effect on the beam, in the case of halfwave plates or passing s-polarised light in the case of polarisers. To locate the home position a two step procedure was used. Firstly the location was coarsely identified by placing a known polarising filter in the beam and monitoring for extinction of the light. The home position was then finely traced by using

a sample with known SHG response. Typically, a sample of de-ionised water was used as this gives zero SHG response at input polarisation 0^o, output polarisation 0^o. A series of experiments would be performed, over a reduced input angle range (for example -5^o to 5^o) to confirm the minimum.

7.3 Electronics

The outputs from the photo-diode, and photo-multiplier tube were processed by a pair of boxcar gates (Standford Research Systems SRS200), time synchronised to the q-switch output from the laser power supply. Data collection was performed on the output of this using a national instruments data acquisition card. A block diagram of the electronics is presented in figure 7.2

7.4 Data collection

Data collection ran under control of the acquisition software described in the next chapter. Data points were generated as the mean measurement of (typically) 1000 laser pulses. These mean values are presented with the statistical standard error calculated for the dataset at the same time as the mean, this is the value plotted as the error-bar on experimental data presented here. Periodically, throughout the collection, the laser source was shuttered and a background measurement made. By shuttering the laser source the signal measured under experiment conditions gives a good representation of the noise in the data due to electrical interference or stray light entering the system. The mean of the nearest-neighbour pair of background points was then subtracted from each data point to reduce the effect of electrical noise and offsets present in the system. The decision to take periodic background measurements was made as it was noticed that there was a small, slow, creep in the background measurements over the length of an experiment.

7.5 Experimental procedures

7.5.1 Glassware cleaning and sample makeup

Unless stated explicitly, all references to water in this section should be assumed to be "high purity" water, as produced by a UV treatment and Reverse Osmosis filter unit.

The SHG experiment is incredibly sensitive to very low concentration samples, hence care must be take to ensure that the glassware and sample handling equipment is clean and free from contamination. To ensure this, all glassware was cleaned with a multi-step

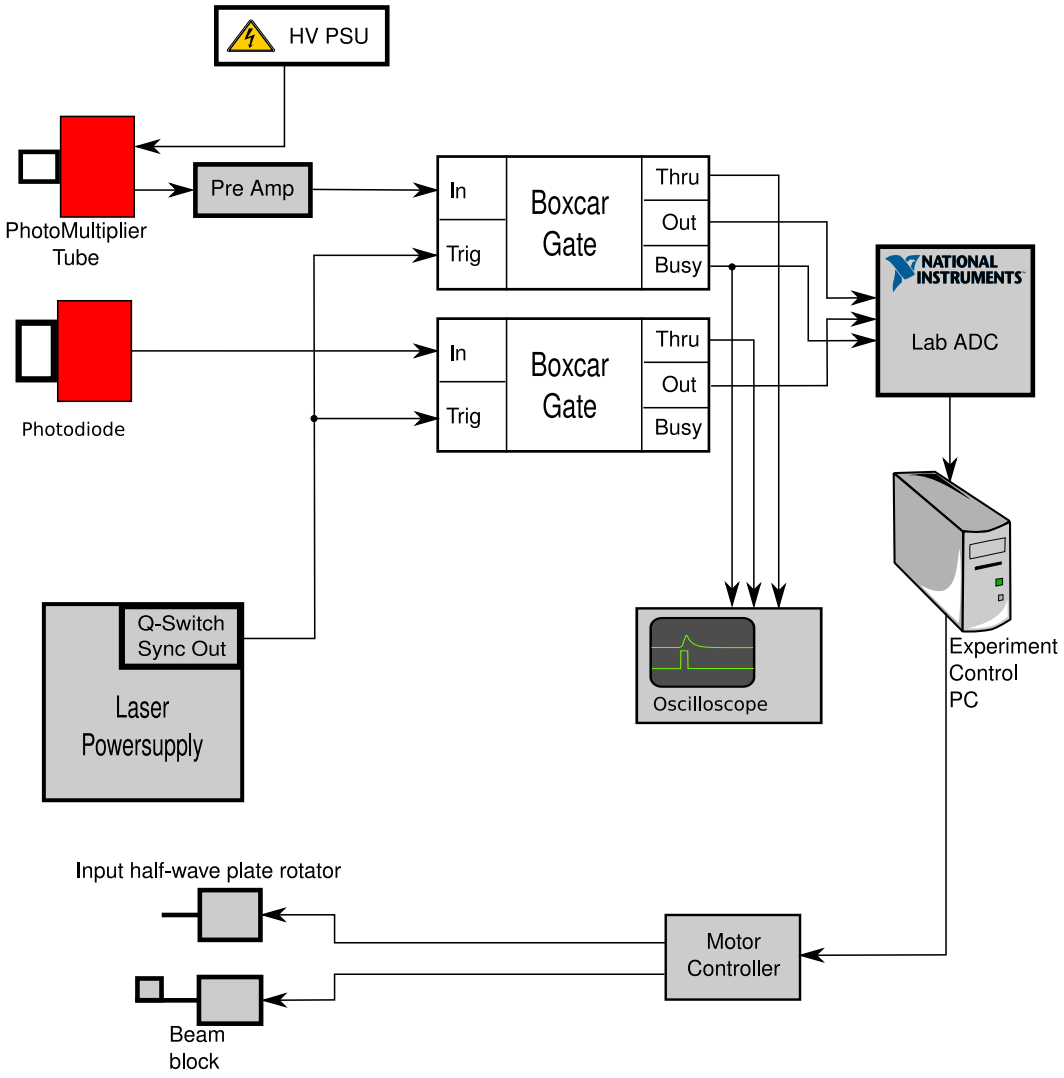


FIGURE 7.2: Block diagram of the electronics in the SHG Rig.
The oscilloscope is disconnected, and the connections terminated at 50Ω *during data capture.*

procedure designed to remove all expected contaminants. This procedure started at the end of an experiment, when the sample holder was rinsed with an appropriate solvent (typically water, acetone or ethanol) to remove the sample before being rinsed further with water.

Before each experiment, all glassware was washed with a dilute detergent solution, then irrigated with excess water. The glassware was then soaked with Nitric Acid to remove any remaining contaminants, before being thoroughly irrigated with water from the same source used for sample preparation. Glassware was then force-dried using a jet of dry nitrogen.

Samples were prepared using glassware cleaned in the way described. Typically a higher concentration stock solution, for example 10^{-4} mol dm^{-3} to 10^{-5} mol dm^{-3}, was first

prepared, before being diluted to the concentration needed for the experiment. All samples were degassed by bubbling with dry nitrogen to reduce the risk of sample degradation by oxidation.

7.5.2 Apparatus Checks

Beam alignment and optics zero point.

Periodically it was necessary to confirm that the optics of the rig were aligned, so as to produce the maximum signal. The rotational position of the polarisation-dependent optics were checked at the same time, to ensure that the experiment was performing as expected.

Beam alignment is a standard optical procedure, in the case of the SHG rig a number of irises were permanently fixed in the beam path, centred on the optimal path. During normal operation, these were left slightly wider than the beam at that point so as to not impede its travel. If alignment changed then it becomes quite obvious, as a green light showing on the leading side of the iris.

The SHG experiment relied on knowing the rotational alignment of two polarisation dependent components, the half wave plate that controls the polarisation of the incoming light, and the polariser that selects the polarisation of light reaching the monochromator. The output polariser was fitted in a graduated, manually operated mount, the alignment of which was fixed, hence all that required checking was the setting at each use (read from the graduations on the mount). The input halfwave plate was mounted in a motorised, computer controlled mount. This mount however had no accurately marked graduations and no electronic zero-point. In use a combination of software malfunction, and mechanical imperfection led to the zero point slipping. To locate the zero a two step process was followed, firstly a coarse location of the zero was found by comparing the light to a known polariser, looking for the point at which the light leaving the polariser extinguished. This gave an approximation of either 0^o or 180^o relative to the wanted zero. By measuring the polarisation SHG signal of a known sample, the zero point was more accurately located. A sample of pure water is known to give no SHG response for a polarisation combination of 0^o in and s-out, and a maximum on a p-out scan is found for an input polarisation of 0^o-in. Small scans of the input polarisation (-5^o to 5^o) were performed for S-out to locate the zero point and a single measurement was then made for 0^o-in p-out to ensure that measurements are taking place a 0^o in, and not 90^o in.

Laser power stability

As is shown in equation 7.6, SHG response is dependent on the intensity of the incident light. For this reason, confidence was needed that the laser source was providing a stable, consistent amount of energy in each pulse; without this the measured response

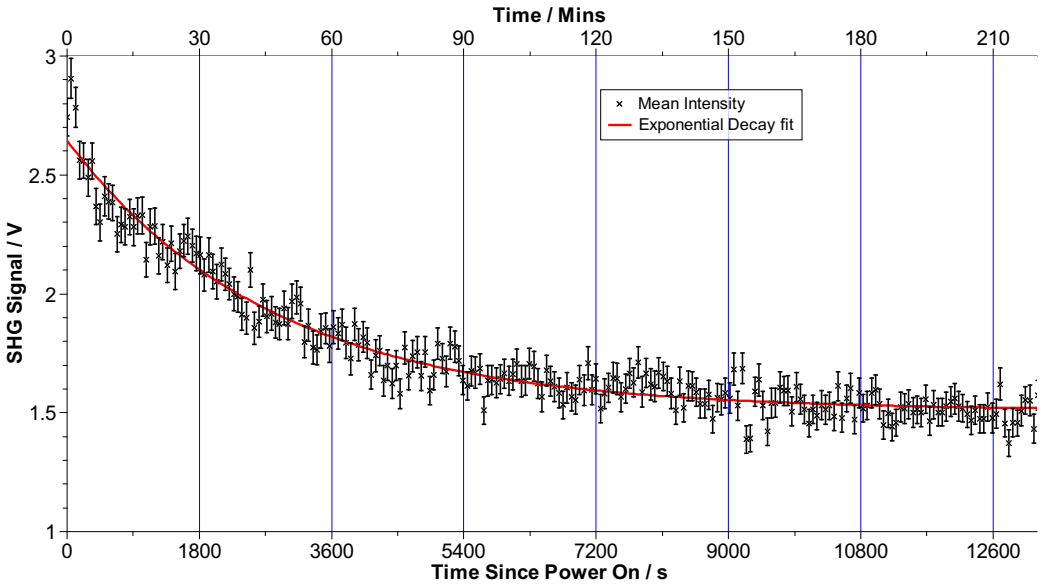


FIGURE 7.3: Laser power stability is checked by measuring power output over an extended time period. Error bars shown are standard error in the mean of the dataset represented by each point on the plot.

would have needed to be corrected for power fluctuation before any other analysis could be carried out. To check the stability of the laser, data were collected over an extended time period (typically many hours) by monitoring the SHG response from a know stable sample (e.g. water), and the power measurement from the reference photo-diode. This was then plotted as signal against time in order to locate a window during which the laser power was stable. This also gave an indication of the state of the laser system, as when it was in good order the power of the output stabilised more quickly, and remained stable for longer. During normal experimental data collection, the apparatus collected data from the photo-diode and plotted these alongside the SHG data. It was important to give this passing attention during every collection, as it was the primary indicator of instabilities in the system, which can then prompt a more thorough investigation.

Figure 7.3 shows a typical warm-up curve for the MiniLite laser. In this case the power is considered "stable" after 150 minutes of warmup, that is the time after which the longer term power change is less than the short time scale jitter.

7.5.3 Compound behaviour

Monochromator Scan

Second harmonic generation produces light at exactly double the frequency of the incident light; in an ideal situation, a plot of signal against wavelength or frequency would

show a sharp peak at the second harmonic and nothing around it. Unfortunately, this ideal situation is rarely found experimentally and a wavelength dependent plot will typically show a non-zero baseline, and may show peaks for light generated by other processes, for example fluorescence effects. Light from external sources which is reasonably constant throughout the experiment will be measured during the background collection (the data collection with the laser source shuttered as mentioned previously), and hence will be removed when this is subtracted.

Light being generated through other effects was more difficult to deal with. In order to improve signal to noise in the SHG signal, the monochromator was usually operated with wider than optimal slit separation. This made it less frequency specific but also gave a higher tolerance in the optical alignment of the apparatus (the target would be a 1 mm or 2 mm wide slit, rather than a 0.5 mm wide slit). If the unwanted radiation isn't actually at the second harmonic frequency, but is nearby, then by reducing the size of the monochromator slits, the accepted frequency band was narrowed.

Following traditional practice, an important step for optimising of the experiment to get best SHG response, was the performance of an experiment scanning the monochromator, usually, for a couple of polarisation combinations. The results from these scans could indicate the presence of other light of similar frequency, and, if it exists, lead to investigations to remove or minimise it. These investigations included a number of other experiments, for example, comparing the monochromator scans at various powers and performing other experiments at non-SHG frequencies to discover the behaviour of the other light (for example to discover if it has a polarisation dependence).

As an example of this, figure 7.4 shows two Rhodamine-6G (R6G) monochromator scans , one for a polarisation combination that produces SHG (red) and one that doesn't (black), showing that in both cases the contribution to signal from non-SHG emissions is similar. For comparison, a water sample measured under similar conditions is also shown. This displays a much weaker SHG peak, with very little signal at other nearby frequencies. It should be noted that the monochromator has a consistent 2 nm offset, as seen in both the red R6G and water curves, hence the peak is centred on 264 nm, not 266 nm when viewing the raw data. This offset has been corrected in the included figures.

Power Response

As shown in equation 7.8, SHG response is related to the square of the power of the incident radiation. Deviation from this behaviour would usually indicate the presence of some other light being produced at or near the second harmonic frequency. Hence, an experiment in which SHG response is measured with respect to the incident power can provide a useful diagnostic when optimising experimental procedure. It should also be remembered that the measured data will contain deviations from the I^2 behaviour, shown in equation 7.8, caused by limitation in the apparatus, most notably at higher

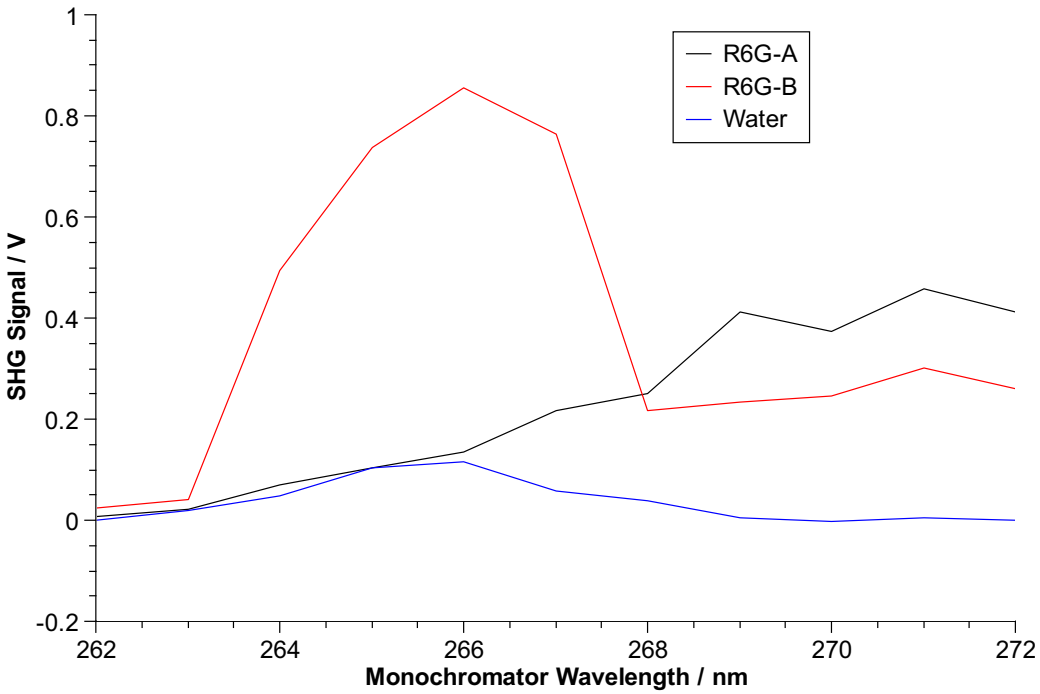


FIGURE 7.4: Comparing Monochromator scans.

and lower powers where there may be too little light to trigger detectors or too much light such that the detectors or measurement electronics become overloaded.

7.5.4 Polarisation Scan

In section 7.1.5 the dependence of SHG signal on polarisation and the relationship to the sample being studied were presented. In order to resolve these parameters experiments were performed at a range of combinations of polarisation, the data from these were fitted against the theoretical response, allowing the molecular parameters to be resolved. For the polarisation experiment, measurements were made across a range of input and output polarisations. Typically, a range of input polarisations was scanned against one output polarisation, this scan was then repeated for further settings of the output polarisation. This is shown diagrammatically in figure 7.5 which should be contrasted with figure 7.6 which shows the flow for an experiment containing a single controlled variable. The sharing of backgrounds is an important point. It both saves time during collection and is used in the post analysis of the data to link sets of data collection together into the experiments that produced them.

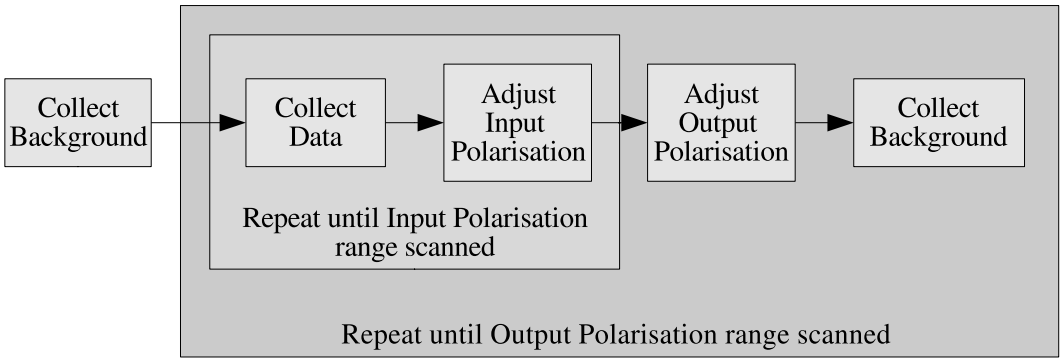


FIGURE 7.5: Process flow for an SHG experiment examining response with polarisation change.

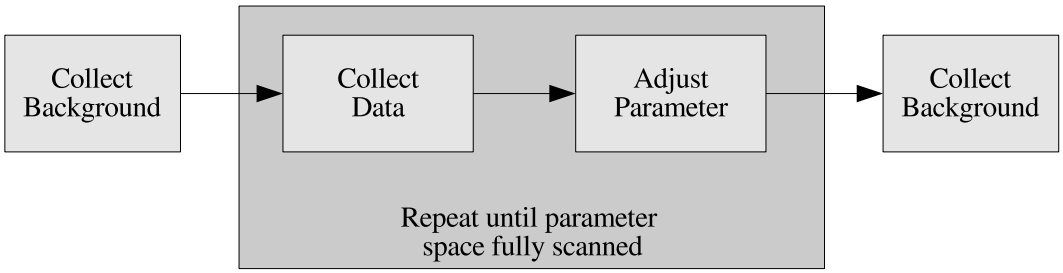


FIGURE 7.6: Process flow for an SHG experiment examining one parameter.

7.5.5 Concentration Isotherm

The SHG response is linked to the interaction of the incident light with molecules at the surface, hence it is reasonable to expect that the measured response will be dependent on the coverage of molecules at the surface (or in the case of liquids the proportion of sample to solvent, where solvent and sample have different responses). It is reasonable to expect a smooth change in the signal reaching some saturation point as the bulk concentration increases, if it is assumed that the packing arrangement of the molecules at the surface doesn't change. If, however, there is a conformational change then the SHG response will change. This may be detectable as a kink or jump in the concentration isotherm. The plot of SHG response against concentration (or better surface coverage) may also indicate concentrations which will need closer investigation, for example regions where the response is more "typical" and hence suited to providing representative measurements of the other diagnostics.

Chapter 8

Second Harmonic Generation
The Data Collection Exercise

8.1 Background

8.1.1 The SHG experiment

The experimental procedures for SHG measurement, as described in chapter 7, involve illuminating a sample with a high power light source (typically a pulsed laser), and measuring the intensity of light leaving the sample whilst adjusting one or more parameters of interest. In the Southampton experiment, a Nd:YAG laser with internal doubling crystal is used to produce green light (532 nm) at a pulse rate of up to 20 Hz. The light is then attenuated to give powers at the sample surface of up to 5 mJ/pulse, although typically lower powers of 1-2 mJ/pulse are used. The apparatus is able to monitor the SHG response with respect to polarisation and power of the incident light for specific output polarisations, at both the SHG and nearby wavelengths. In the past, it has been used to monitor variations in these factors with respect to changes in temperature[71]. By changing or modifying the sample, comparisons with respect to concentration and composition can also be made.

To improve the repeatability of the data collected, normal practice is to collect measurements for a set of laser pulses, with the experimental parameters fixed, and to take the mean of these. Periodically, a background measurement is made by shuttering the laser output and recording data corresponding to the laser firing as it would for data collection. Typically, only one parameter is varied at a time, with a background measurement being taken periodically throughout. If two parameters are being varied, as is the case for polarisation scans, then a nested loop arrangement is typically used where the first parameter is varied as described for a single parameter, but after the second background is collected the second parameter is changed and another set of the first

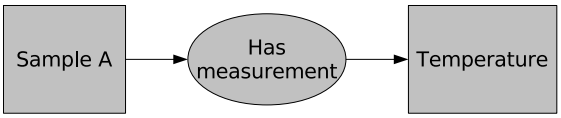


FIGURE 8.1: A simple 'triple' relationship.

parameter collected. This can then be repeated until the complete space of the second variable has been covered.

The experiment's software needs to collect experimental parameters and meta-data from the user, configure the apparatus and then perform the data collection loops; traversing the experiment space as defined by the parameters entered by the user. As the program progresses, it should be storing this information and the raw data into a database. It also needs to alert the user to the current condition of the experiment, and to request user intervention as needed (for example to move some optic which hasn't been automated). Once collection has been completed, the software should trigger the automatic analysis of the data.

8.1.2 Data Storage

There is a need to store experimental results and associated meta-data. At the start of this work two options were considered, the use of either a triple store or a relational database. At the time, triple stores were a new technology, with the promise of delivering flexibility and extensibility meaning that they can cope with datasets that evolve without needing to completely rewrite the storage schema when they do. However, to achieve this flexibility there is a scalability trade-off which was starting to become apparent.

Data within a triple store is stored as the relation between objects and values, for example the object "sample A" is related to value "temperature" by the relationship "has measurement". This is easier to consider diagrammatically as shown in figure 8.1. "temperature" is then related to value "21" by the relationship "has value", "temperature" is also related to "centigrade" by "has units" as shown in figure 8.2. Other measurements can then be added as shown in figure 8.3. As this shows, adding another measurement is simply a case of reusing the same tools, and can be used for almost any measurement without adding to the underlying tool-set; with a relational database the schema would need to be extended, and values found to fill the new fields for all existing samples. For a combination of sample and measurement that hasn't been collected the triples aren't included, comparatively with a relational database a value would need to be stored in that field to indicate that the measurement is missing.

The SHG data have a rigid structure which is repeatable across experiments, hence is suited to storage in a relational database; it is therefore possible to design a schema

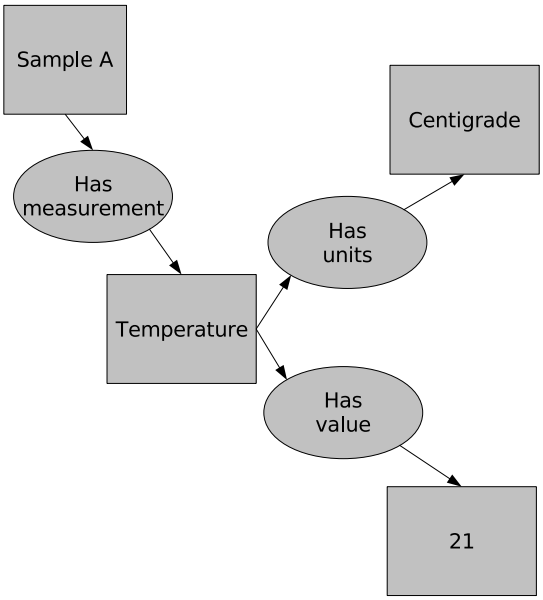


FIGURE 8.2: Triples to store a single measurement.

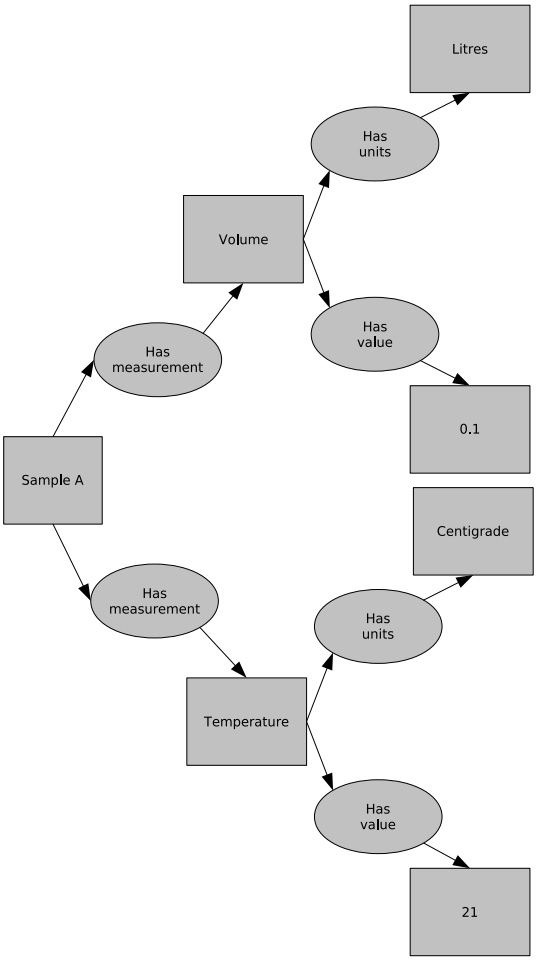


FIGURE 8.3: Extending the triples to further measurements.

that will be stable for all experiments. In this case, a relational database was preferred, as the scalability has already been proven and the flexibility of a triple store was not required. Design of the database is addressed in section 8.2.2.

8.1.3 Data Publication

Once data are stored in an ordered, well structured, database they become more useful by being easier to locate and recall. However, data constrained within an institution are effectively lost, greater benefit is only gained upon publication. This work considers the publication of SHG data for a number of purposes each of which may require a slightly different interface. These uses can be internal, for example when sharing the data with colleagues or when reviewing old data with new assumptions, or external, when data is published in the literature or shared with a wider audience.

Data collaboration with colleagues requires the sharing of data and processed results, followed by discussion of the conclusions that can be reached. Without the aid of collaborative systems, this may take the form of circulating a spreadsheet by email followed by an email discussion, or perhaps the sending of data printouts through the post ahead of a group meeting. With the application of e-science technologies, this process can be shortened and made more efficient. Recently, the Internet has seen a huge growth in collaborative tools initially for the creation of web-pages and subsequently as a collaboration tool within the workplace context. Content management systems based around web forms are one of the early examples of this, followed closely by forums and discussion boards. Both technologies allow the user to enter text (either plain or with formatting markup) into a form on a web-page which is then used as part (or all) of the content for another page. These technologies extended into the current trend for Wikis (Collaborative websites that anyone can edit) and Blogs (a contraction of Web Log - the term coined to describe the keeping and publishing of journals on the web). In the blog concept, the reader is being shown the current thoughts and opinions of the writer presented in a chronological order, possibly with comments from other readers. Work by Bleecker[72] describes the concept of objects (or equipment) that write blogs, creating the term blogject. This follows the same mentality of moving from the journal to the blog - taking a previously private log, and placing it on the web. Within the scientific community, this concept has the potential to be very useful – once on the web the data can be made available to a much wider audience, for them to both use and comment upon.

Following from a requirements gathering exercise, it was discovered that the **data analysis** associated with polarisation scanning SHG experiments was (initially at least) quite repetitive, attempting to fit the dataset to a theoretical model and then comparing the results. Performing this process manually is quite time consuming, and hence was normally only performed for data sets that "appeared promising" on the initial raw plot.

By automating this fit procedure, all datasets can be fitted to the model and the data compared – whilst a "poor" dataset may not fit exactly, if its fit parameters approximate those of a "good" dataset, then confidence in all the data improves. If a "poor" fit is later discarded, this discarding is then based on a more thorough understanding and can be better justified. Such a tool must maintain links to the raw datasets, as this is the provenance chain that allows later review of the conclusions that may be published based on the analysed data. This goes hand in hand with the collection and storage of environmental and other data associated with an experiment. When a complete picture of the experiment space can be presented, any conclusions become far more justifiable.

When **publishing** results in an academic paper, the focus is usually on the conclusions drawn from the data and experiments, rather than the dataset itself. This is in part due to the premium on space in the printed work. In recent times, it has become common for publishers to create an online archive of "supporting" data that can be viewed alongside the printed work both to justify the conclusions, and also to allow the reader a deeper understanding of the subject. The content of this archive is totally at the whim of the author, who can include as much or as little data as they see fit. More useful would be the inclusion of all the data that supports a publication; a system that either produces an archive to be handed over to the publisher or that gives access to the data at the author's institution would make it far easier for all the data to be released to the reader. At the same time, the author should also be ensuring that there is an adequate record of the apparatus used to generate the data, preferably in the form of reviewed papers focusing on the instrumentation or at least by making available un-reviewed documents describing the instrument.

Previously ignored has been the possibility for an apparatus to publish reports, or mini-papers, whenever an experiment is performed, and that these papers might be placed into the public domain as valid research in their own right. By making these papers available as they are captured, an alternative way of considering the publication of supporting data is provided.

Such autonomous publication would generate a huge number of additional papers for consideration, far more than peer review could cope with; but without the review process, how can the reader have confidence in the data contained within the paper? Confidence could be increased by providing the reader with the provenance of the data being viewed in the form of access to the raw data, apparatus design details and reference to previous published results from the same apparatus or author. A question that would still need to be considered concerns the judgement of the academic reader. If the assumption was made that the data and provenance published were authentic, then would the reader's judgement be adequate to negate peer review? Equally, how could the authenticity of the data be proven?

8.1.4 Data Review

In chapter 3, the concept of publishing all data, and the increased value of being able to recall data, is discussed. It may enable the application of new data analysis techniques, or simply allow the scientist to review how their data has changed over the longer time frame; they maybe suspicious of a slow degradation in some part of the apparatus, for example. By storing all the experimental data in a database, this process should be much easier, as all the data should be instantly available, and ready to be included within an analysis. However, for this to be the case, there needs to be some human interface to the database. For the experienced computer scientist, this may simply be the management console of the database engine or an abstracted interface in their favoured programming language. However, many physical scientists would desire a lower level of complexity, finding it easier to proceed with the cumbersome task of manually dealing with a large number of serial-numbered data files, referenced from their paper lab-book. By careful application of some e-science technologies, a compromise situation can be reached, which provides the physical scientist with the power of the computer scientists abstract interfaces, but through an easily managed set of tools.

Scientists tend to think of their data in terms of when it was collected, what the sample was, and what the purpose of the data collection was. An e-science tool should respect this, allowing them to filter the data presented by familiar headings (for example by selecting dates, samples, and lists of controlled variables), and should present the result of this search quickly, and in a straightforward fashion. A scientist would be used to dealing with tabulated data in a spreadsheet, so presenting a tabulated summary of their database search would be a logical conclusion, but by also providing graphical previews of the data (either pre-rendered and recalled from the database, or generated on the fly) then confidence that the correct data has been found could be made even higher, especially considering that when using the raw data files alternative, one of the early steps in checking that the expected data had been found would be to load it into an appropriate viewer and generate a plot.

8.2 Implementation

8.2.1 Data acquisition system

The acquisition system was built in Visual Basic using National Instruments Measurement Studio modules to access the capture card, and the Microsoft ADODB framework for database access. All other functionality was either coded manually, or built using the standard VB modules. Data collection flows through a nested loop as shown in figure 8.4. This was primarily designed for the polarisation experiments where scanning took place across two variables, input and output polarisation, but it is easily adapted

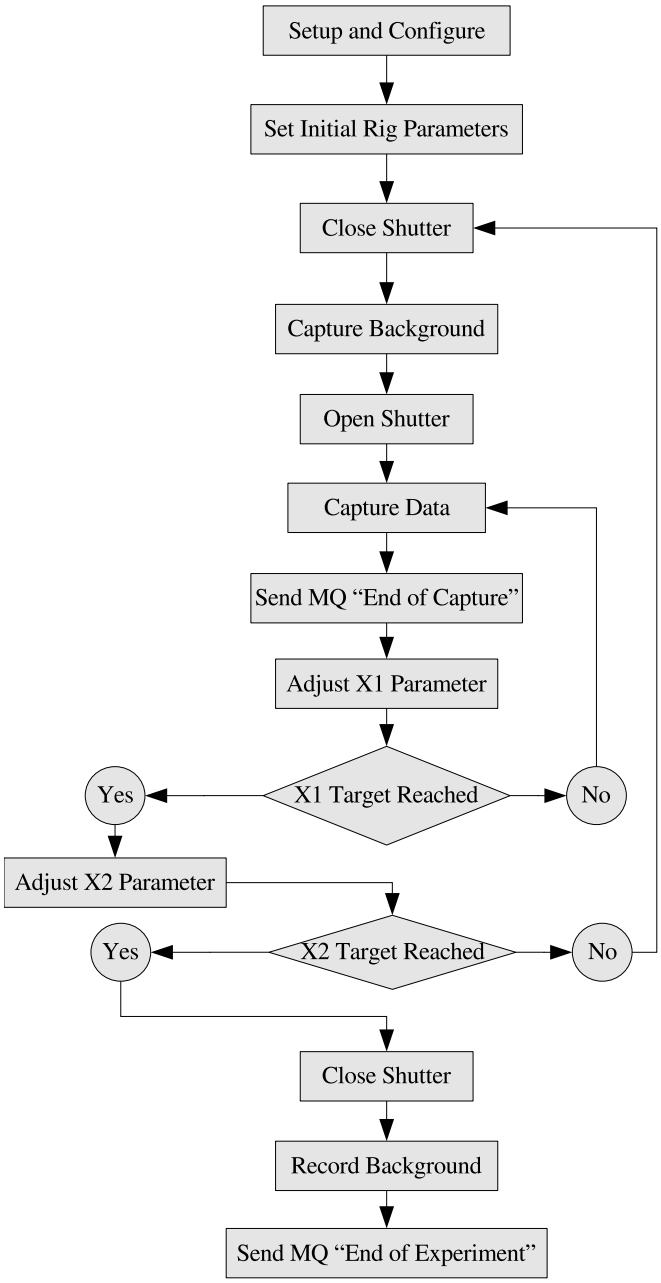


FIGURE 8.4: Process flow within the experiment data capture software.

to single variable scans through inclusion of a "null" handler to negate the outer loop (the inner loop then being used for the variable under investigation). The two loops progress by scanning sequentially across the two experimental parameters indicated at the experiment design stage.

The software starts by prompting for user authentication (Figure 8.5), before presenting a meta-data collection and experimental design interface (Figure 8.6). The experimental design consists of setting parameters of the system and experiment to be carried out,

including the sequential progression through a range of values for two of those parameters (the inner and outer loops). The interface also collects other parameters of the experiment (typically settings from the rig electronics) so that these are stored with the experimental data (Figure 8.7). The user is finally prompted to enter file details (for the local backup stored on the capture PC) before collection starts. During data collection, the previous raw data set and a summary of progression of the current data collection is presented on screen (Figure 8.8).

Throughout data collection, the software publishes status updates to the laboratory message broker. This allows for the remote monitoring of experiment progress, and also the generation of triggers for other experiment dependent services. One example of this is the automatic analysis system discussed in section 8.2.4; when an experiment ends an automatic fitting of the experiment data to a theoretical model occurs, and the result published into an electronic log. The use of the message broker in the SHG laboratory was introduced in chapter 4.

8.2.2 Data Storage

To support the acquisition and processing of data from the SHG experiment within an *e*science framework, it was required that a central data-store existed to hold the raw data from the experiment. It was decided to provide this using a relational database, as the parameters involved are reasonably static, and hence generating a rigid schema was possible, as discussed in section 8.1.2.

Design of the database schema was completed collaboratively with other researchers involved in the database's use, both experimentally and from the point of view of producing software [20]. It was a requirement that the database produced would deal with all the current experiments.

The initial step in creation of this database schema was to define, in broad terms, the data to be stored. This extended from the raw data collected by the acquisition card, through the experimental parameters, to information about the sample being studied and the chemist performing the experiment. These data were then re-classified by how often they changed (rarely/per sample/per dataset etc) as shown in table 8.1. By separating the fields into logical groups (e.g. data about users, data about chemicals, data about samples etc) and the splitting these groups according to the "frequency of change" (table 8.1), the schema shown in figure 8.9 was generated.

This was implemented using postgreSQL. PostgreSQL was chosen as it was a readily available, open source database engine which supported foreign key enforcement and transaction control. MySQL was also considered, but at the time had no support for foreign key enforcement or transactions.

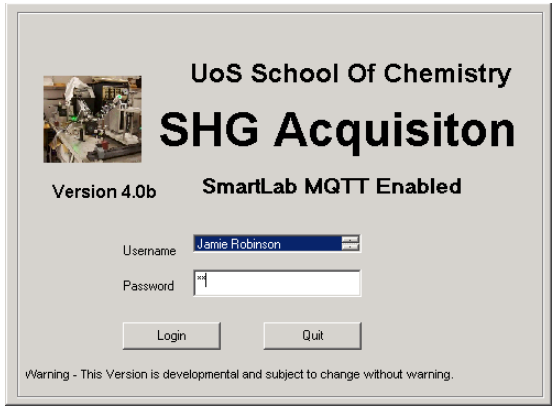


FIGURE 8.5: SHG Software - Login Screen
Username/Password authorisation is used to identify the operator

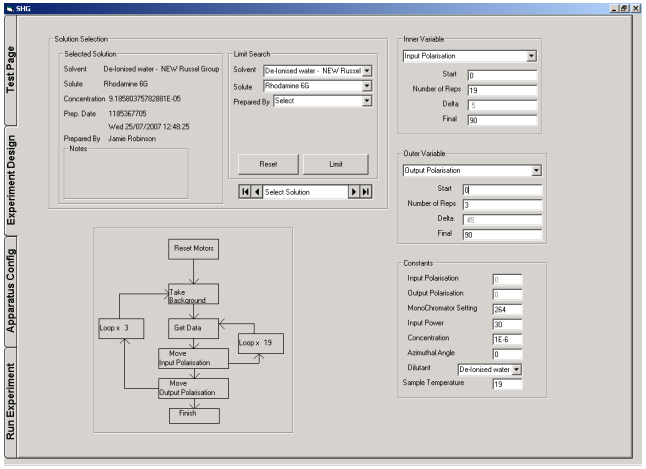


FIGURE 8.6: SHG Software - Experiment Design
The operator is prompted to enter the ranges for the parameters under investigation, and the fixed values for the rest

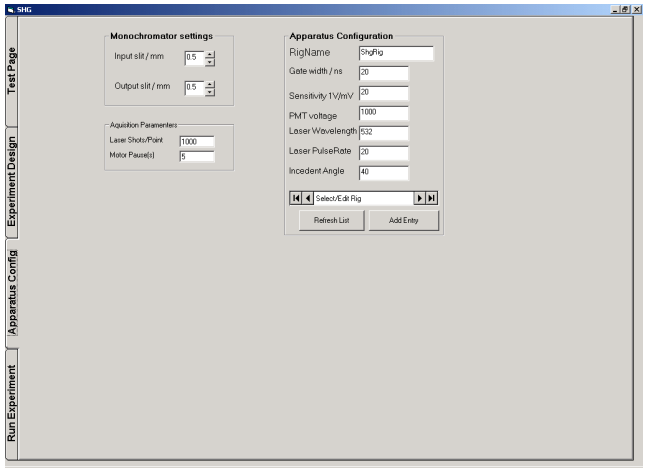


FIGURE 8.7: SHG Software - Meta-Data Entry
Values for various other parameters on the apparatus are recorded, as they may be important for later analysis

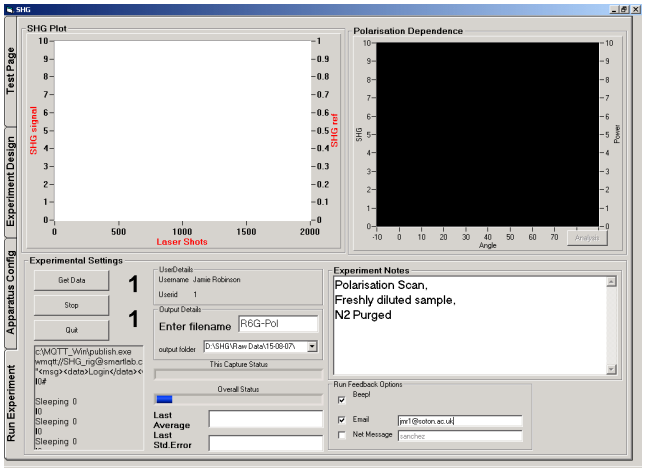


FIGURE 8.8: SHG Software - Data Collection
Whilst data are being collected, the operator is given feedback through progress bars, and plots of the the raw data

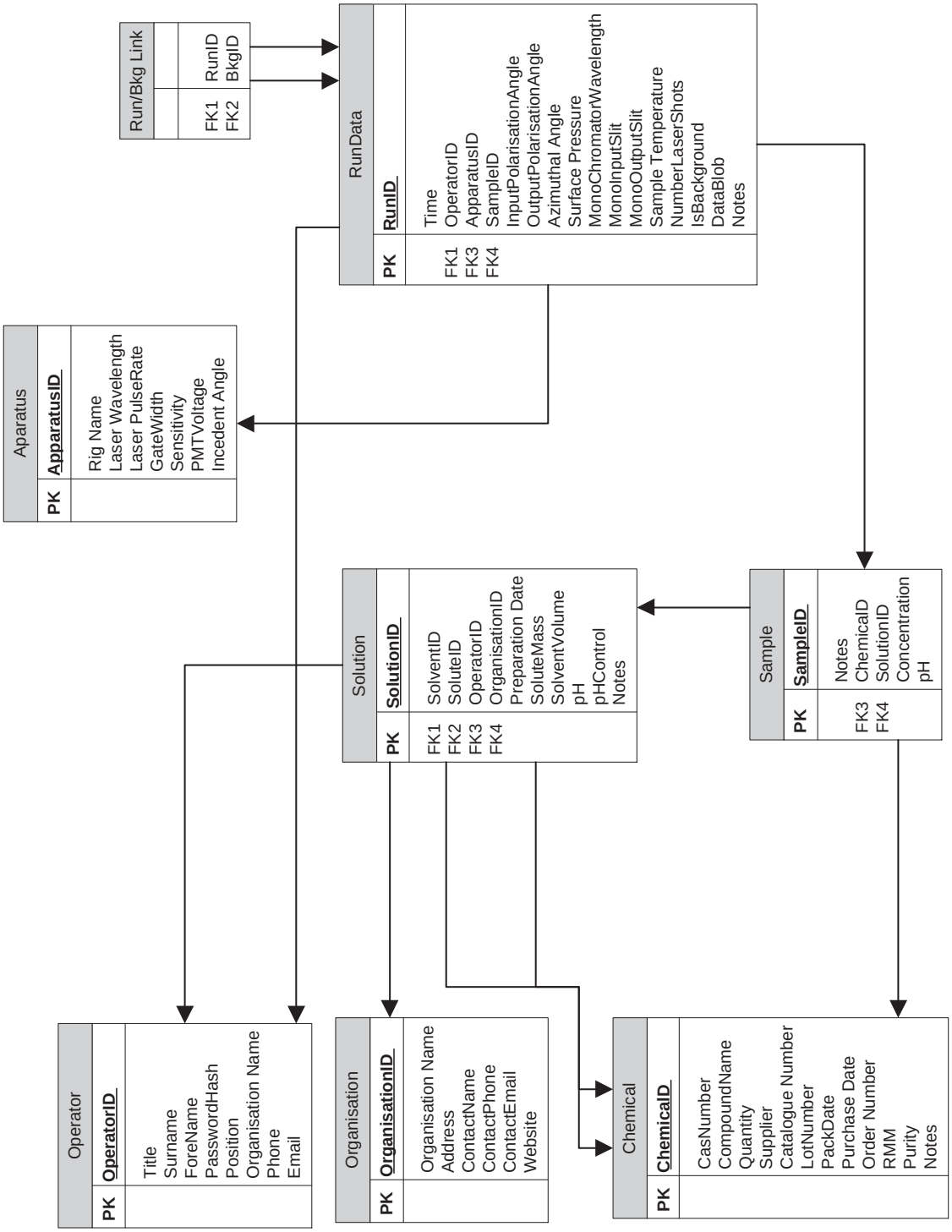


FIGURE 8.9: Relational Database Scheme for storage of SHG related data.

Frequency of Change	Rarely	Per Sample	Per Experiment	Per Run
Value	Operator Details *(name, address, etc)*	Solvent	Sample	Polarisation
	Stock Compound Details *(name, supplier, etc)*	Solute	Operator	Laser Power
	Organisation Details *(name, address, etc)*	Concentration	Apparatus parameters *(PMT Voltage, Gate settings)*	Monochromator Settings

TABLE 8.1: Separating the SHG data fields according to their frequency of change.

In order to populate this database, two routes have been used. Firstly, data were inserted directly by the acquisition system as they were recorded. Secondly, data from previous experiment were recovered from raw data files and added to the database. This retrospective insertion proved to be necessary as it allowed the development of other software that relied on the repository, before adequate experimental time had elapsed. Fortunately, the previous data included much of the meta-data needed in the raw data files, the remaining meta-data was recovered from lab-books.

8.2.3 Publication @ Source

As part of work with Hong-Chen Fu, the SHG database formed the back-end for a published data recall tool.

The comb*e*chem *e*paper[20], written during the course of this thesis, described and implemented an infrastructure that gave the reader of an academic paper the ability to recall the underlying data directly from the publishing institution. In this example, the paper was an investigation into the surface behaviour of aqueous solutions of benzo-crown-15 using surface tension, SHG and empirical calculation. The SHG results presented in this work linked back to data in the database described here. The paper consisted of two main parts; one discussing the technology of publication, and the second showing the example investigation of benzo-crown-15 solution surface. The second part contained interactive graphs and tables that linked to pages providing a summary of the raw SHG data used for that graph point or table entry. This linking from within graphs was made possible using an SVG graphing utility developed at the University of Southampton for the purpose. Graphs within the paper were generated once, and embedded into the electronic version. The summary pages, however, were designed to be generated on request. At the time, the technology used wasn't sufficiently developed

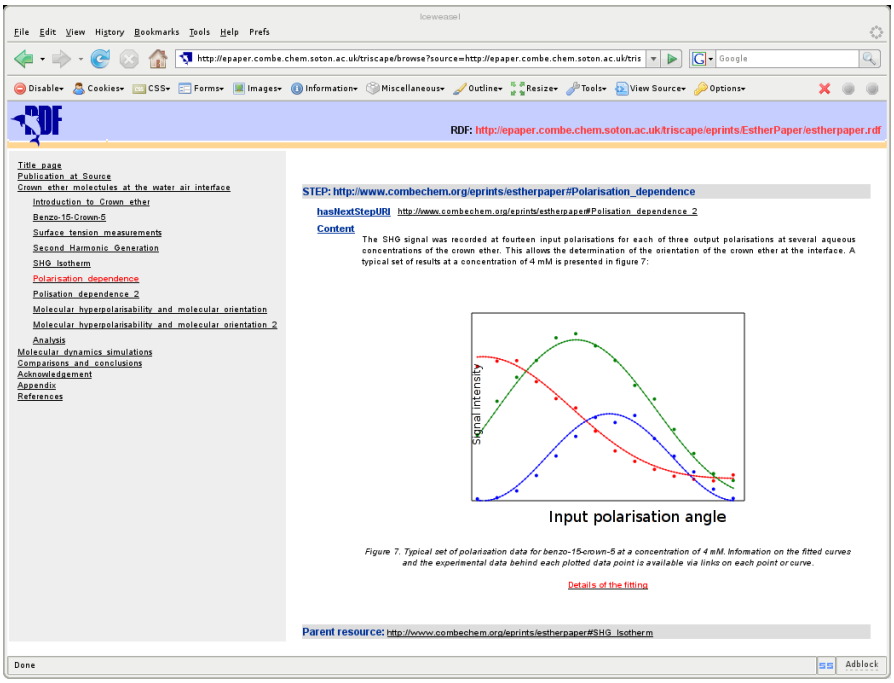


FIGURE 8.10: The triscape browser, developed for the comb*e*chem project. The RDF encoding of the paper includes the section structure displayed as the left navigation and links to other media, in this case an interactive SVG graph.

to generate the interactive features completely; a certain amount of manual crafting was needed during the authoring process, to tag the text and data appropriately. This tagging was needed to add context to the paper's sections and to the data points. By providing more automation of, and tools for, the data analysis of SHG, this manual crafting can be reduced as the tools become able to record much of the context required as it is generated.

8.2.4 Automatic publication

One of the comb*e*chem project goals was to improve the ease with which chemists perform research, and hence enhance their efficiency. An important trend in modern science is that of collaborative working, which can be achieved in a familiar, cross platform fashion through the use of web-technologies. Andrew Milsted, a researcher on the CombeChem project, was involved in investigating the use of "web logs" or "blogs" within the laboratory space[58]. That technology fits well with the aims of this work and so it was decided to integrate the SHG data into it. The chemistry blog tool provided the same hypertext publishing tools that would be found in any traditional blog, but has the added facility to link data files to blog entries, and provides a web-based tool to display them.

In order to get data from the SHG experiment into the blog with the minimum of user intervention, an agent was written that monitors the experiment status topic on the laboratory message broker, waiting for the 'end of experiment' notification. This notification includes a list of database references for all the data-points in that experiment. On receiving this message, the agent recalls the experiment data from the database. It then locates and recalls the background measurements for each data point, and subtracts these to give final values. Checks on the data are then performed to ascertain what type of experiment has just run (for example, polarisation scan, monochromator scan or a stability check). Based on the type of experiment, the agent generates a gnuplot script to produce an appropriate plot; this is then executed. An entry is published into the blog containing the background subtracted averages file, a PNG graphic of the plot, and the experiment's meta-data, as extracted from the database. Figure 8.11 shows this process diagrammatically.

Software was produced that allows for the manual injection of "experiments" into the blog. The purpose of this is to allow the inclusion of old experimental results into the blog and to allow the combination of discrete runs into new "experiments". For example, a sequence of runs may be carried out as multiple experiments on the apparatus, but it may be useful to include them as a collection in the blog, alternatively an experiment may be run over a much wider range of variables than needed, a subset of these could be selected for inclusion in the blog. All "experiments" added to the blog in this way are flagged to show that they were generated by manual selection rather than as the result of an actual experiment. A typical blog entry is shown in figure 8.12

8.2.5 Experimental data recall and analysis

The database structure used for raw data was very convenient for machine reading and processing, but was difficult for a human operator to follow manually. To make best use of such data, a method of searching and extracting results is required. Where a theoretical model for the system's behaviour exists, part of the analysis is likely to involve fitting the experimental data to the theoretical model and there are obvious benefits to automating this process. Additionally, data of this nature are often displayed graphically, so any recall system should include graph generation capabilities. All these factors were taken into account in the design and implementation of the architecture for the system which is summarised in figure 8.13.

The review system extracted data from the experimental database, and then performed a background subtraction. It then grouped these data together into "experiments". A plot and summary data file was generated based on this grouping. Where a suitable model exists, a fit can be performed and the model parameters presented. The model for polarisation dependent SHG experiments presented in section 7.1.5 was implemented to demonstrate this. This whole procedure was scripted, and the results stored in a

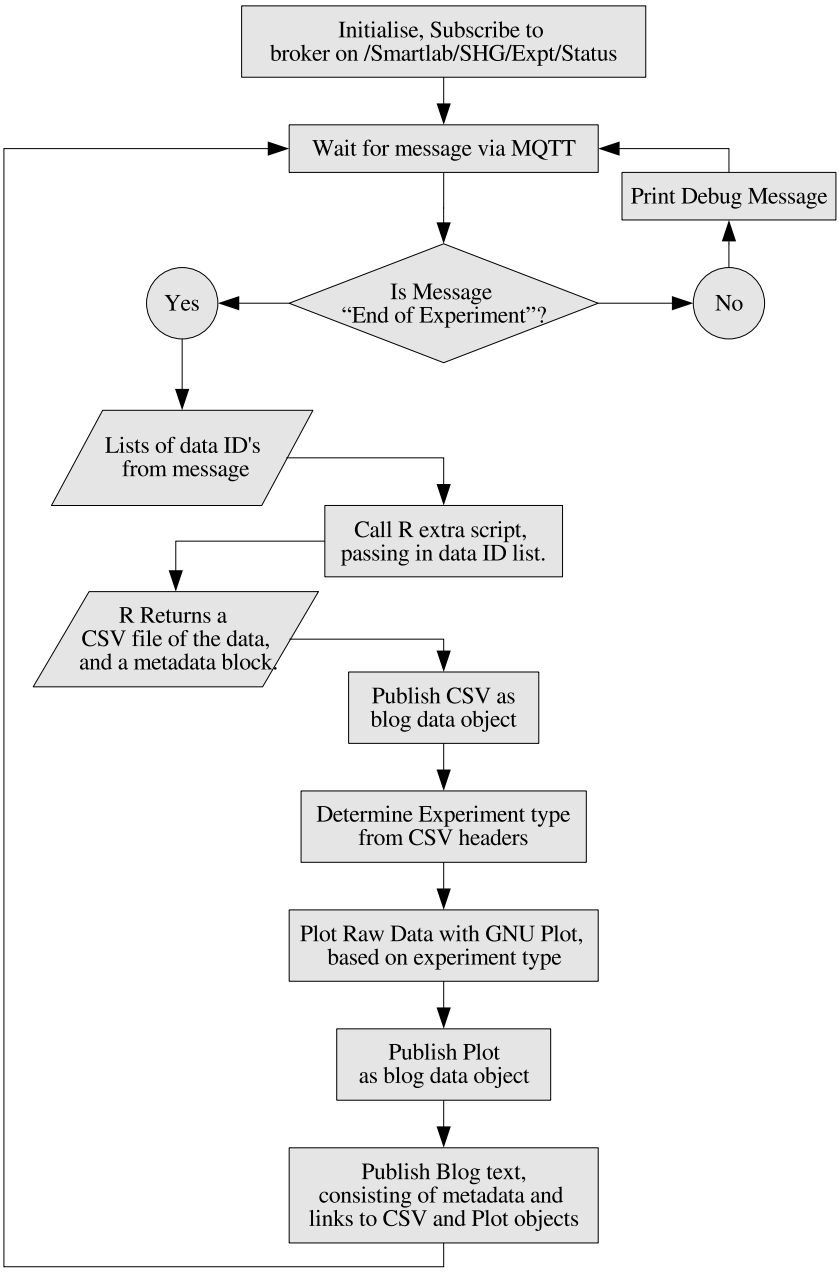


FIGURE 8.11: Process flow within the auto-blog agent.

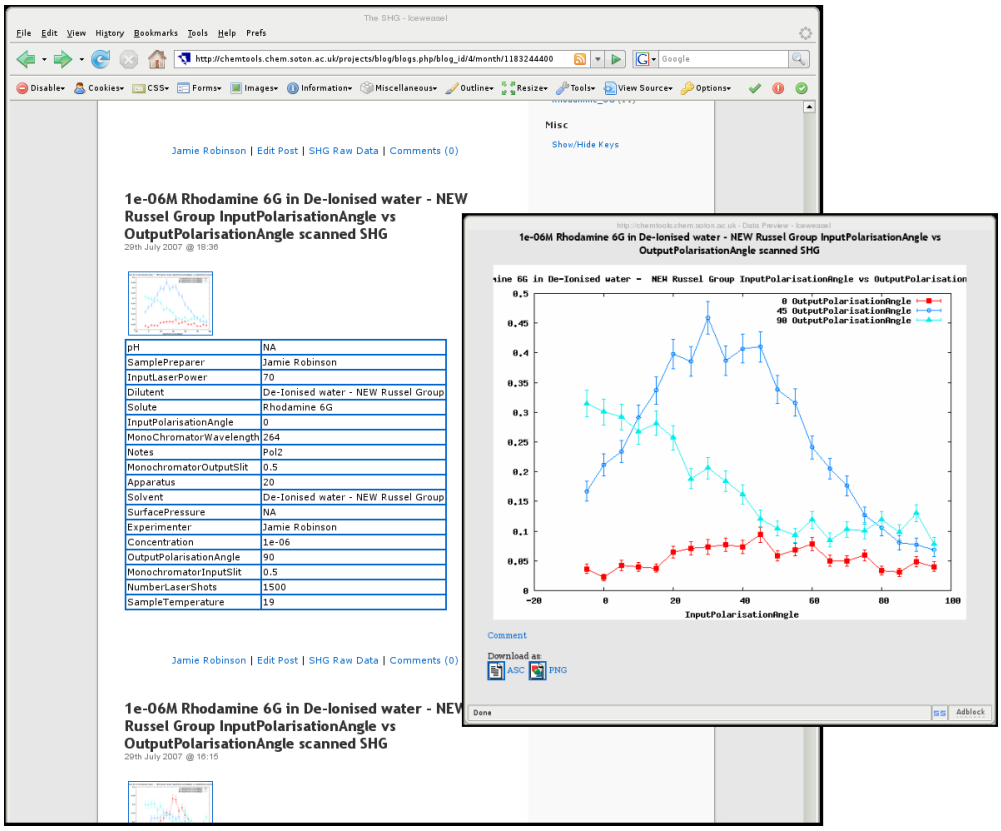


FIGURE 8.12: A typical SHG blog entry, in this case showing polarisation data for a Rhodamine 6G sample. Data in the blog can be downloaded both as graphics and as raw data.

database, the contents of which are presented to the user through a web interface. This extraction and summary procedure used is presented in the next section. This script runs autonomously, and detects when an analysis has already occurred and skips it.

The web interface presented the user with a list of all the experiments in the database which could be filtered by compound, solvent, date and experiments performed (figure 8.14). When the user selected an experiment, a summary of that experiment, listings of the raw data contained in that experiment, and details of all the fits performed on the data set were displayed(figure 8.15). The system also presented the details of any fitting procedures used and any summary calculations performed. The summary page also presented the ability to download the data. From the experiment listing, the user was able to select a group of experiments for which they can then download a tabular summary containing details of the experiment and the parameters of the fits performed (figure 8.16).

The experiment summary database stores the additional data derived during the recall procedure, leaving the raw data to be accessed directly from the original database. The combination of these two data-sources provides a powerful tool from which summaries and other analysis can be performed. The data storage within the summary database is

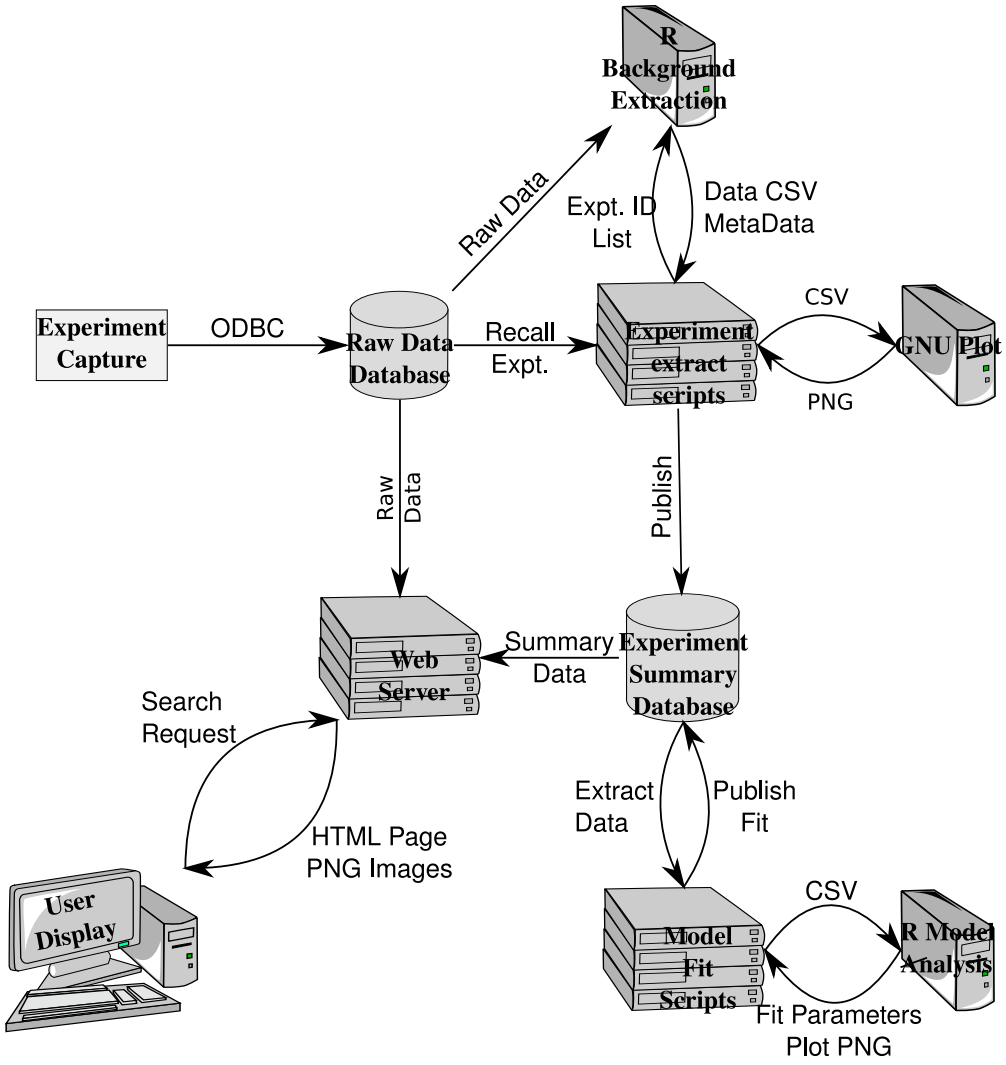


FIGURE 8.13: Overview of the experiment detail recall and extraction system.

sufficiently free-form that it can be used to store any data; data is stored as binary blobs which are referenced to the experiment that they refer to by an experiment ID. This experiment ID then references the rawdata relevant to that experiment. The database schema for the summary database is presented in figure 8.17.

8.2.6 Grouping raw data into experiments

Both the auto-blogger, and experiment summary database refer to groups of raw data as experiments. The auto-blogger benefits from the rig or the user providing the list of raw-data ID's that make up an experiment. The experiment summary/recall tool, however, needed to be able to derive this grouping from the data in the raw data tables. This was achieved based on the assumption that background readings are shared within an experiment and that discrete experiments have their own background readings. The

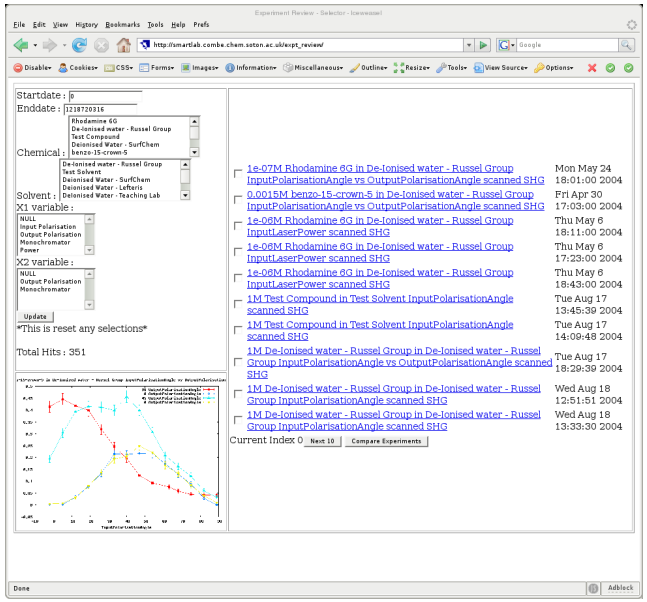


FIGURE 8.14: SHG Experiment Review - Experiment Selection
Experiments in the database are listed chronologically, with the ability to filter irrelevant experiments

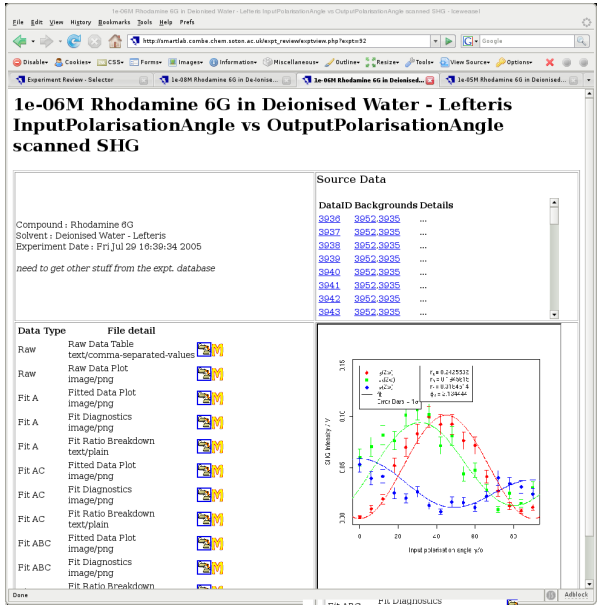


FIGURE 8.15: SHG Experiment Review - Experiment Detail Display
When an experiment is selected, details of the raw data and fitted data can be listed, viewed and downloaded.

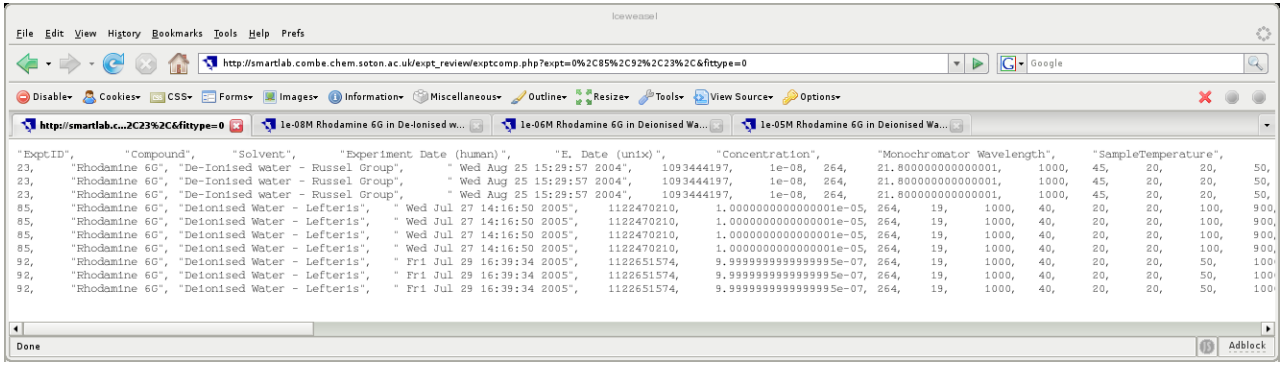


FIGURE 8.16: SHG Experiment Review - Experiment Comparison
Groups of experiment can be selected and a tabular comparison of them provided.

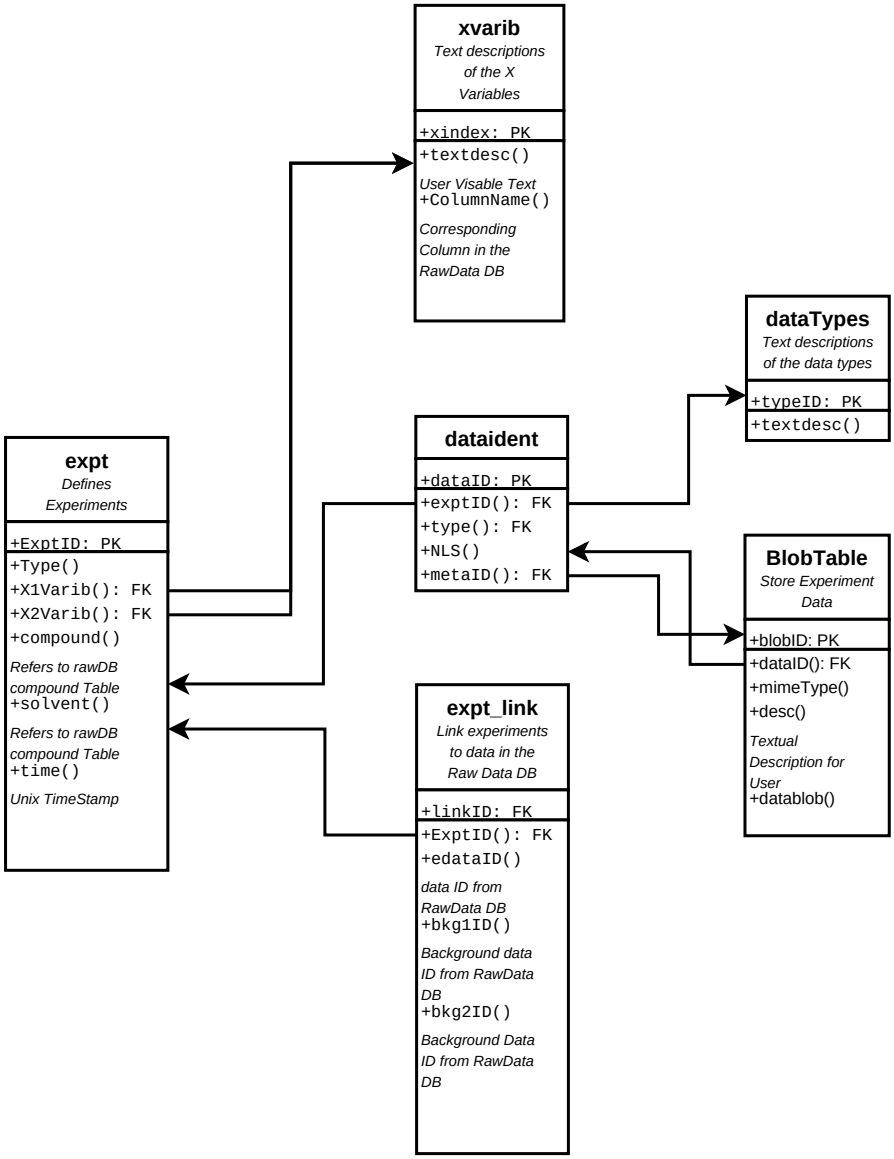


FIGURE 8.17: Relational Database Scheme for storage of SHG summary data.

procedure followed by the tool that builds the summary database is summarised in figure 8.18. This process traverses over all the data points in the raw-data database and, where the current point isn't already referenced in the summary database, begins the experiment build and extract procedure. The list of data points in an experiment is built based on the assumption that background measurements are only shared within an experiment; hence, by recursively selecting all the data associated with a list of backgrounds, then all the backgrounds associated with those data points, the whole experiment can rapidly be found. The list of data points can then be passed to the extraction procedure (which is common with the auto-blogger).

The raw data extraction script was written in the R language, and is presented diagrammatically in figure 8.19. This recalls all the data associated with the data id list

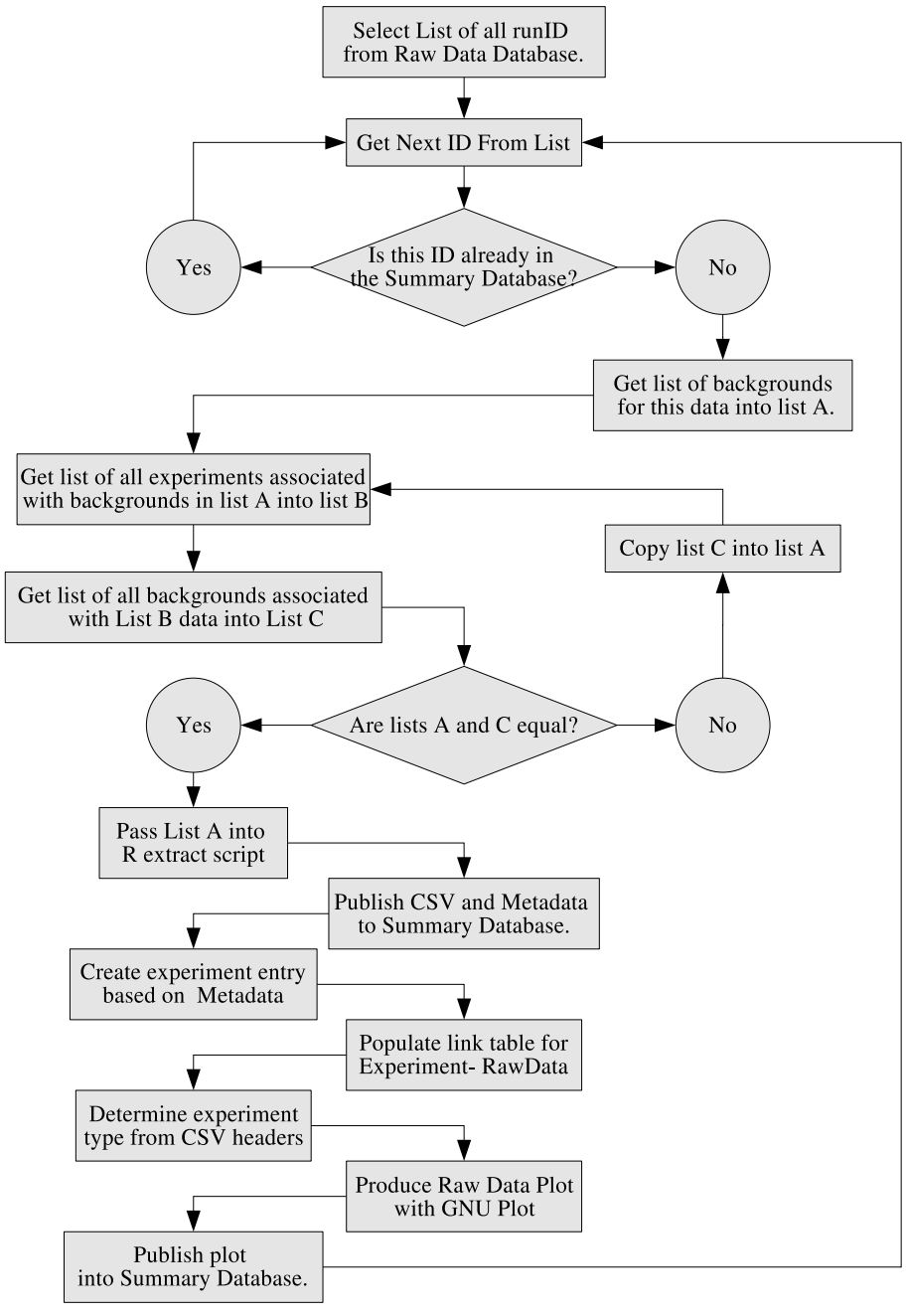


FIGURE 8.18: Process used to group raw data into experiments.

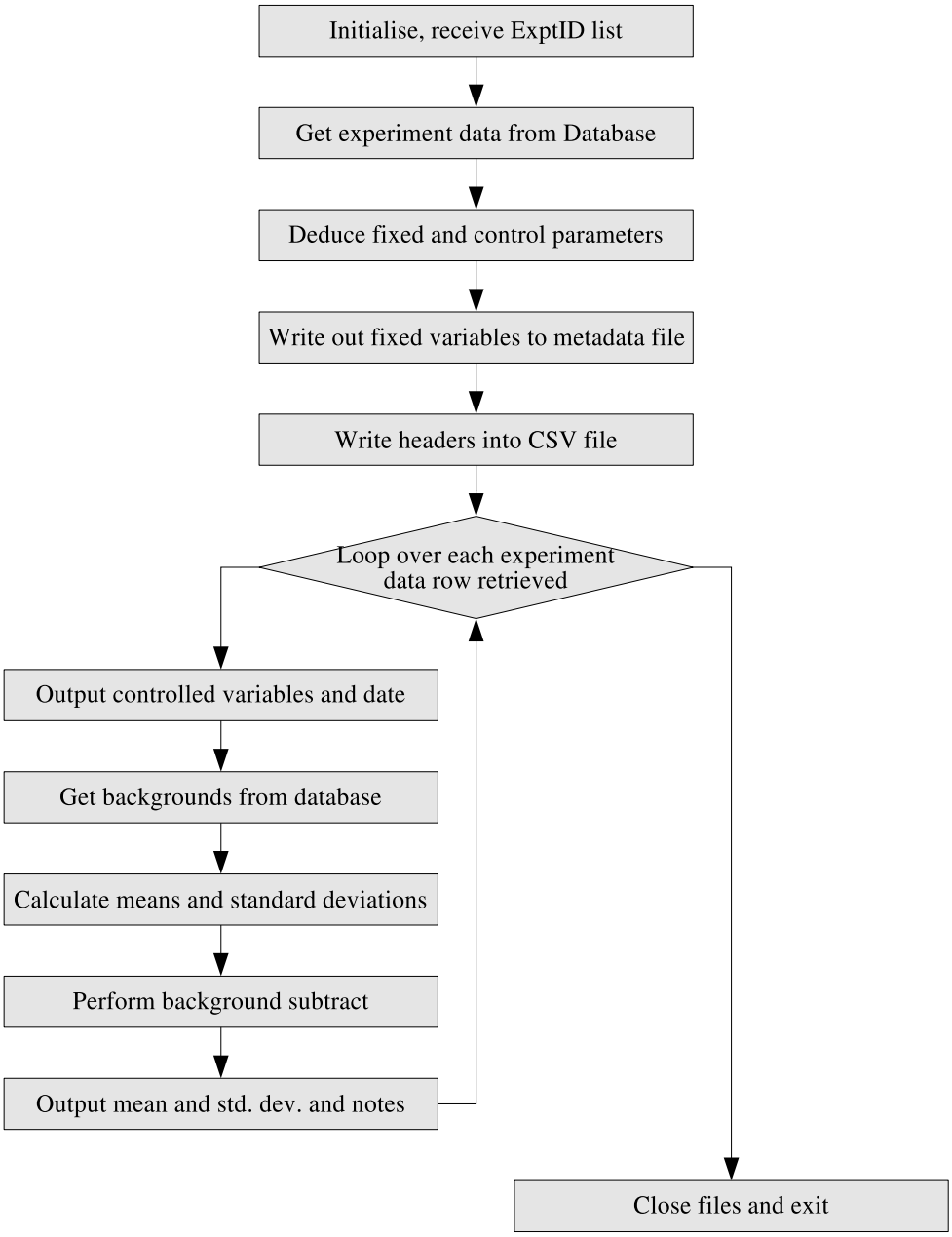


FIGURE 8.19: Process used to translate a list of experiment ID's into a human readable summary.

passed to it and then, for each id, recalls the associated background measurements. It performs a background subtraction before calculating the mean and standard error for that point. Once this has been completed for all points, it returns a summary of ID, controlled variables, mean and standard error as a comma separated data file, and all the fixed variables as a variable list. These files are then processed by the parent program; in the case of the auto blogger they are published as blog data files, while the summary database build tool places them into the summary database.

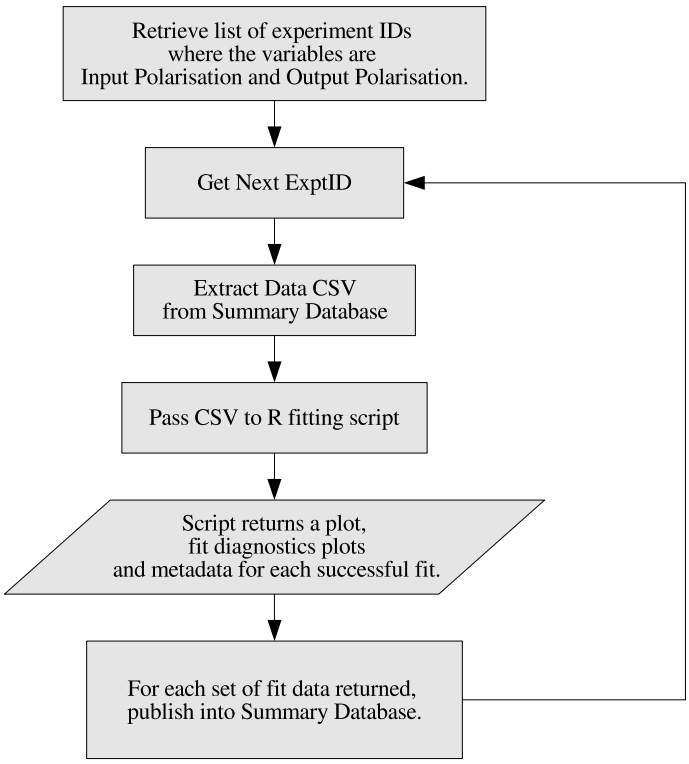


FIGURE 8.20: Process flow for the automatic fitting of polarisation scanning SHG experiments to an existing model for storage with the summary.

Once the raw data are available in the summary database, those experiments for which a theoretical model exists should be fitted, and the fit parameters presented for comparison. This was implemented for polarisation experiments using the procedure shown in figure 8.20. The analyser tool recalls a list of all experiments for which input and output polarisation are the controlled variables. It then traverses this list one experiment at a time, in each case passing the summary CSV file to an extension of the R analysis script produced by Welsh et Al in their 2004 paper[73]. The extension to the script causes it to fit for all combinations of real and complex parameters, collecting all the plots and parameters. These are then all published into the summary database. All these scripts are included on the attached data CD, and are documented in appendix B.

Chapter 9

Rhodamine 6G

Rhodamine dyes are members of the Fluorene dye family, chosen for use as a laser gain medium due to their strong absorption maximum in the green light region. Rhodamine GD (also know as R6G, Rhodamine 590 and Basic Rhodamine Yellow, and shown in figure 9.1) is particularly popular because of the close proximity of an adsorption maximum (530 nm) to that of the second harmonic of Nd:YAG laser sources (532 nm). R6G lases in the range of 555 nm to 585 nm, with a maximum at 566 nm. Previous studies of aqueous R6G solutions have show it to form dimers, and bigger aggregates, in bulk solution [74, 75, 76]; this is shown to reduce quantum yield affecting fluorescence and lasing effect. It is believed that only the monomer exhibits a lasing effect, and that the aggregates strongly adsorb both pump and emitted radiation[76]. Bulk studies have shown that below a concentration of 10^{-6} mol dm^{-3}, 99 % of R6G molecules are in the monomer form and that at a concentration of 4×10^{-4} mol dm^{-3} 50 % of the molecules are in dimer form.

Understanding of the behaviour of R6G at the surface is far more limited. The surface specificity of SHG allows for surface behaviour to be studied across a range of concentrations, independent of the bulk. Some limited studies of R6G using SHG measurements have been performed; in his PhD Thesis Danos[66] presented preliminary work that

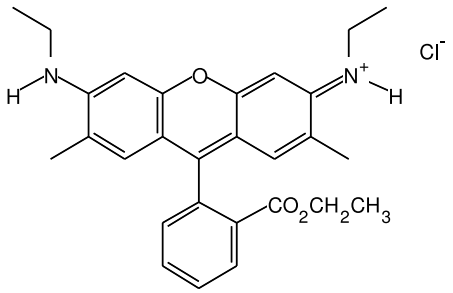


FIGURE 9.1: The molecular structure of 9-(2-(ethoxycarbonyl)phenyl)-3,6-bis(ethylamino)-2,7-dimethylxanthylium chloride (Rhodamine 6G)

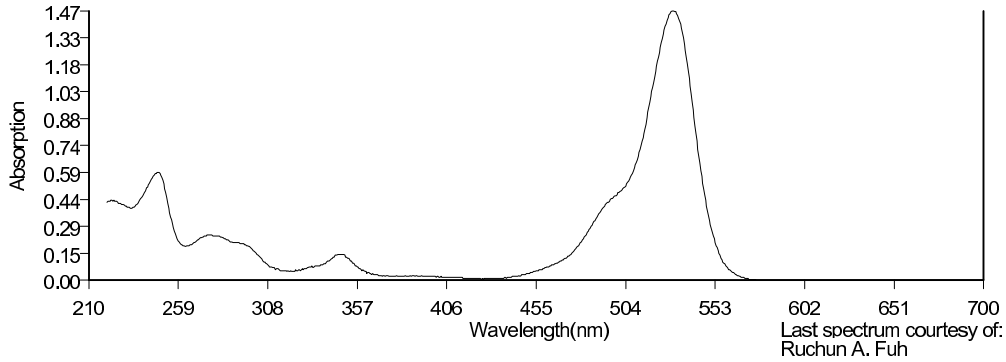


FIGURE 9.2: Published UV-Vis spectra for R6G in ethanol solution.
From PhotoChemCad[77]

showed a change in the SHG response at concentrations above 75 μmol dm^{-3}. He believed this to be due to the formation of dimers at the surface. That work was continued here with the aim of confirming those findings and gaining a complete understanding of the surface behaviour.

9.1 Previous Work

9.1.1 UV-Visible Absorption spectroscopy

Figure 9.2 presents typical adsorption data for R6G in Ethanol solution. It shows clear peaks corresponding to the $S_0 \longrightarrow S_1$ transition at $\lambda = 529$ nm and the $S_0 \longrightarrow S_2$ transition at $\lambda = 347$ nm. It also shows adsorption below $\lambda = 300$ nm corresponding to the lone aromatic ring.

Work was performed to examine the two-photon fluorescence behaviour of many dyes, and the xanthenes were shown to not follow the theoretically predicted intensity squared model as other dyes do with the fluorescence yield at higher powers being much less than expected. A recent study by Fischer, Cremer and Stelzer [78] summarises previous work and extends it, examining behaviour using a femtosecond pulsed titanium-sapphire laser. It is shown therein that study has been concentrated on the λ =750 nm to λ =1024 nm pump region, which leads to fluorescence with a peak around λ =560 nm from aqueous R6G solutions. There is little or no work considering pumping in this region for emission at ultra violet wavelengths. It was found, as a side effect of this SHG study, that pumping at λ =532 nm yields fluorescence around λ =300 nm and extending down to λ =266 nm, the SHG frequency in this study. This fluorescence made it more difficult to isolate the SHG emission as shown in section 9.2.2.

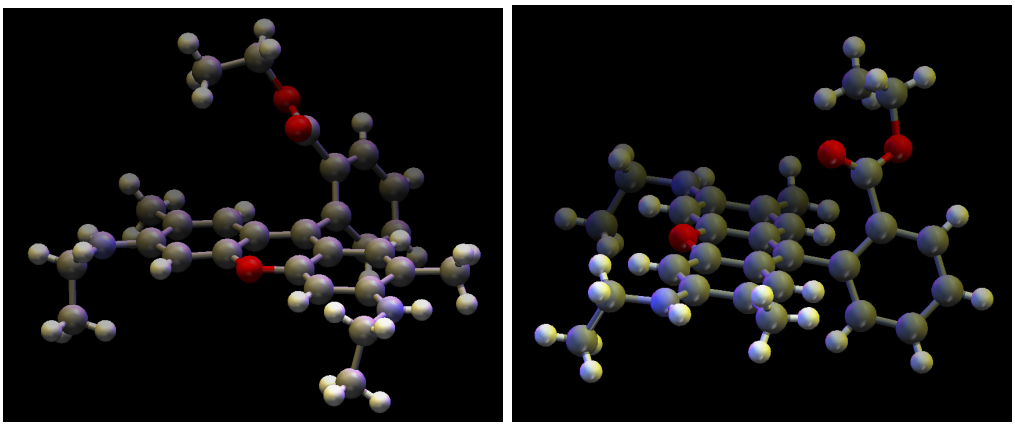


FIGURE 9.3: Crystal structure of R6G.
From XRay diffractometry[79]

9.1.2 Crystal Structure

The published crystal structure[79], as logged in the Cambridge Crystallography Database (CCDC), shows R6G in the solid state to behave much as would be expected, with the xanthene core being forced planar due to its aromaticity and the lone benzene ring having no shared aromaticity with the xanthene core hence sitting perpendicular to it, figure 9.3. It also shows the ethylamino groups sitting together, with the ethoxycarbonyl group falling to the opposite side of the main structure. In aqueous solution studies the crystal structure is interesting as it gives some initial insight to the behaviour and size of R6G molecules, and the way in which dimers may form. Not shown in the figures are the chloride counter ions and water molecules that exist in the mono-hydrate form. Measurements from the crystal structure put the amino nitrogen atoms 10.89 Åapart, and distance from the phenyl C4 to a cord between the terminal amino C as 10.01 Å. Hence, when later calculating coverage values, these should imply a molecular area in the region of 1.1×10^{-18} m^2.

9.1.3 Surface Tension

As part of their study of the photo-ionisation of Rhodamine dyes, Seno *et al*[80] performed measurements of the surface tension of R6G/water solutions in order to derive the surface coverage of dye in their samples. Their data show R6G to have a very low change in surface tension with concentration, especially when compared with the other Rhodamine dyes studied.

It is possible to fit a Langmuir isotherm to the data presented, as described in section 2.2, and this is presented in figure 9.4

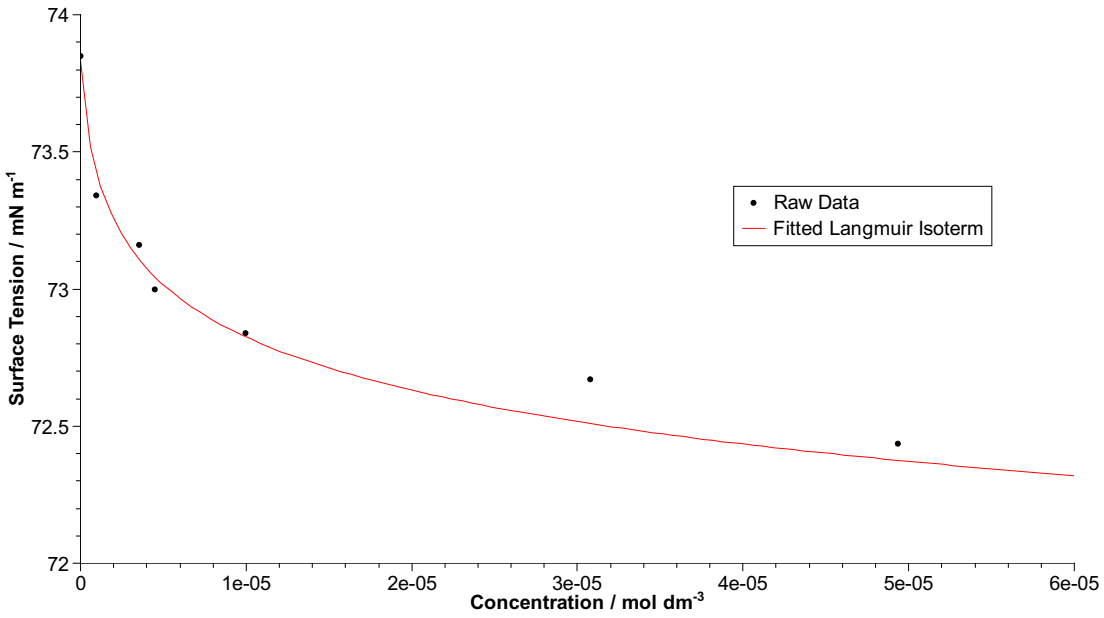


FIGURE 9.4: Surface Tension of Aqueous R6G solution.
Extracted from the Seno et al. [80]. Data fitted to equation 2.36.

9.2 Experimental

9.2.1 Ellipsometry

Ellipsometry measurements were made under the direction of Dr R. Greef, using a Uvisel Ellipsometer from Horiba JobinYvon, which utilises a Xenon arc lamp and has a detector range from 250 nm to 800 nm. Normal calibration procedures for the monochromator, and straight-through optical alignment were used and it was ensured that the apparatus base was level, as a liquid surface is inherently level so the apparatus must be level to avoid related alignment errors. The signal reflected from liquid surfaces is low, hence the signal was averaged over relatively long periods of 0.5 s to 1 s. The majority of the noise in the system was believed to be due vibrations of the liquid surface, hence precautions were taken to avoid these; this included redirecting air flow from the room air conditioners and avoiding personnel movement in the room during data collection. Samples and glassware were prepared using the regime described in chapter 7 for SHG experiments.

9.2.2 Second Harmonic

Second harmonic experiments were performed as described in chapter 7. Rhodamine 6G purchased from Acros Organics was used as supplied and water from a Sartorius Arium-UV reverse osmosis system (resistivity 18.2 MΩ cm, total organic content less than 2 ppb) was used for glassware cleaning and sample preparation. To enhance confidence in the results, a number of experiments were performed to check repeatability and that measurements were within the bounds of the apparatus.

Intensity dependence of the second harmonic signal

In equation 7.8, it is shown that the intensity of the second harmonic signal is dependant on the square of the intensity of the incident radiation. As a diagnostic to check that SHG was being measured, a plot of the log of signal intensity against the log of incident power was expected to produce a good linear fit, with a slope equal to 2 for regions where the apparatus was measuring the second harmonic reliably. Deviations from this form can be caused by other sources of light at the second harmonic frequency and by overloading (or other non-linearity) of the measurement electronics.

Figure 9.5 shows examples of log-log plots for a 10^{-6} mol dm^{-3} R6G solution and for water. It shows a very good agreement for the R6G solution across the whole range of powers measured. The water signal deviates somewhat from the expected fit; this is in part due to the relatively weak SHG response produced by water. If only the water data at higher powers (when the signal is stronger) are considered, then a fit closer to the expected slope is obtained.

Scanned monochromator experiments

Some samples (such as Rhodamine 6G) produce secondary emissions at wavelengths near the SHG frequency, and it is necessary to have confidence that measurements of the second harmonic were not distorted by these other emissions or, if they could not be isolated, that the contribution to the measured signal could be taken into account. Figure 7.4 (chapter 7) shows two examples of R6G monochromator scans, one with SHG emission and one without; these scans were generated using two sets of polarisation settings, one which produces SHG, and one which does not. Figure 7.4 shows that there is little difference in the secondary emission at the example polarisations chosen, and that to estimate the contribution to SHG measurement from the secondary emission, the signal measured for a non-SHG producing polarisation setting is a good approximation.

Figure 9.6 shows data collected from a weak R6G solution. If shows that once the monochromator slits have been narrowed, the SHG response is clearly separable from the other emissions, and that there should be no need for further numerical treatment. In this case, the 'wide slits' setting corresponds to a slit 1 mm wide, and the 'narrow slit' setting to 0.5 mm wide. At the 0.5 mm setting more care had to be taken with

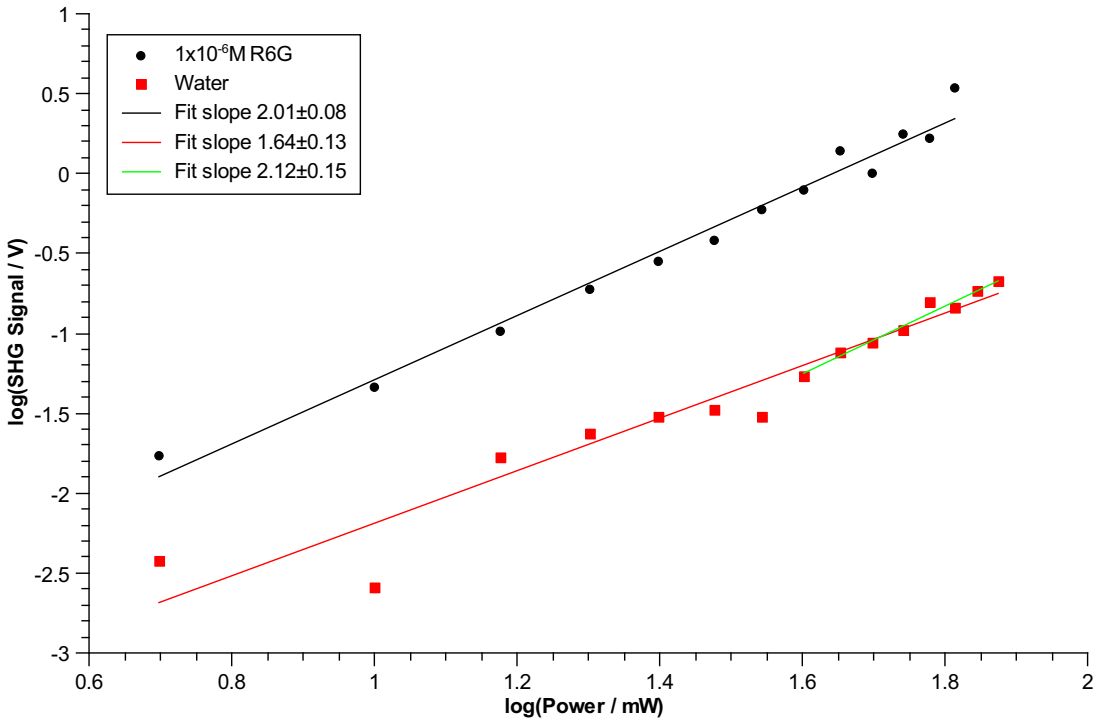


FIGURE 9.5: SHG signal power dependence for R6G(aq) solutions.

the optical alignment of the apparatus, as the laser beam waist was filling the width of the slit separation, hence any horizontal deviation from this could have led to light not reaching the monochromator grating, and so being lost.

Figure 9.7 shows data collected for a much stronger R6G solution (1×10^{-5} mol dm^{-3}). In this case it is much harder to reduce the effect of the secondary emissions on the SHG measurements. To obtain the separation shown in the 'narrow slits' measurements, the monochromator slits were set for 0.25 mm spacing. This resulted in the slit separation being narrower than the beam waist as it passed them, and consequently some signal was lost. With care it was possible to focus the beam down to a tighter spot, minimising this loss. There is, however, a limit to how small the beam can be made, and it was preferred not to do this as the risk of damaging the gratings during alignment is then much higher. The problems caused by surface stability are also increased as the monochromator slit size decreases (ripples on the liquid surface deflect the beam); if the beam target is smaller, then a smaller deflection is needed to stop the beam hitting that target. The data collected indicate that, with care, data collected under these conditions are acceptable; with the restricted monochromator slit size, the contribution from secondary emissions was less that 10 % of the measured signal.

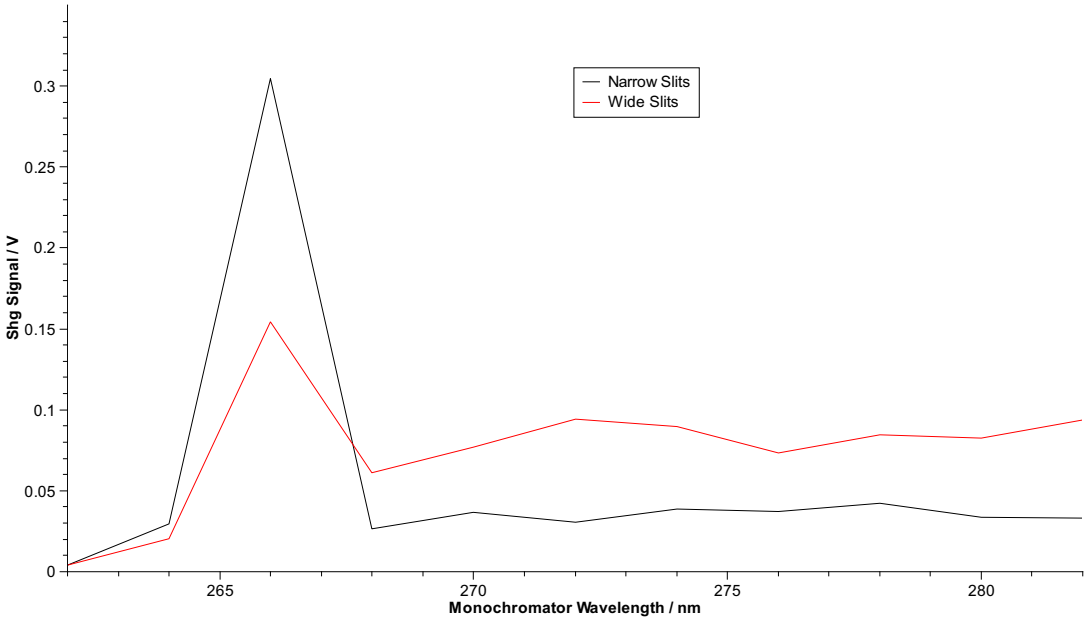


FIGURE 9.6: Monochromator scans for a 10^{-7} mol dm^{-}3 R6G solution.

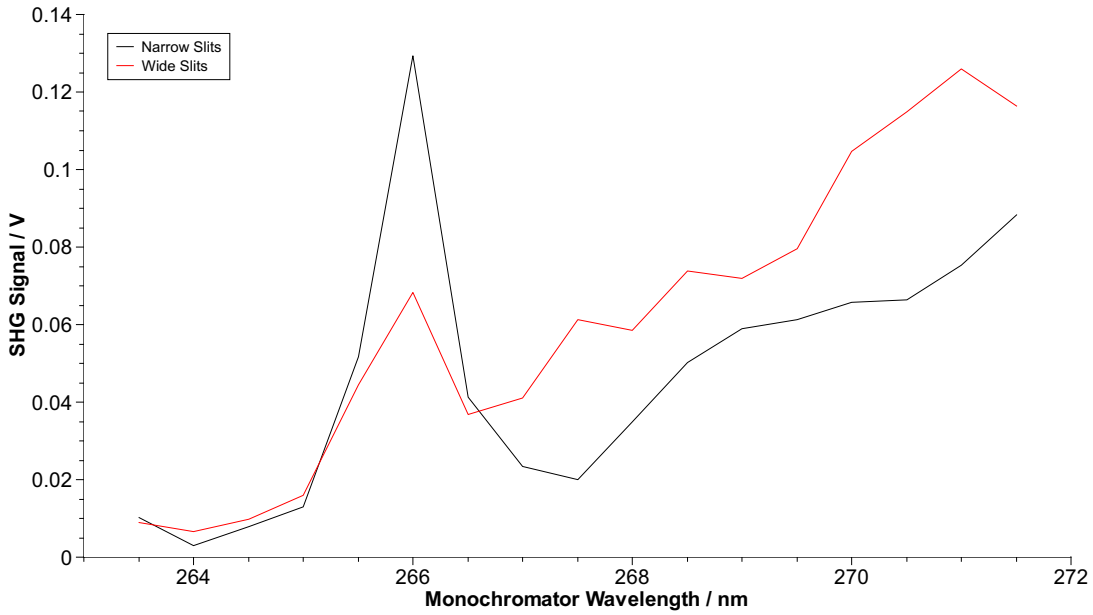


FIGURE 9.7: Monochromator scans for 10^{-5} mol dm^{-3} R6G solution.

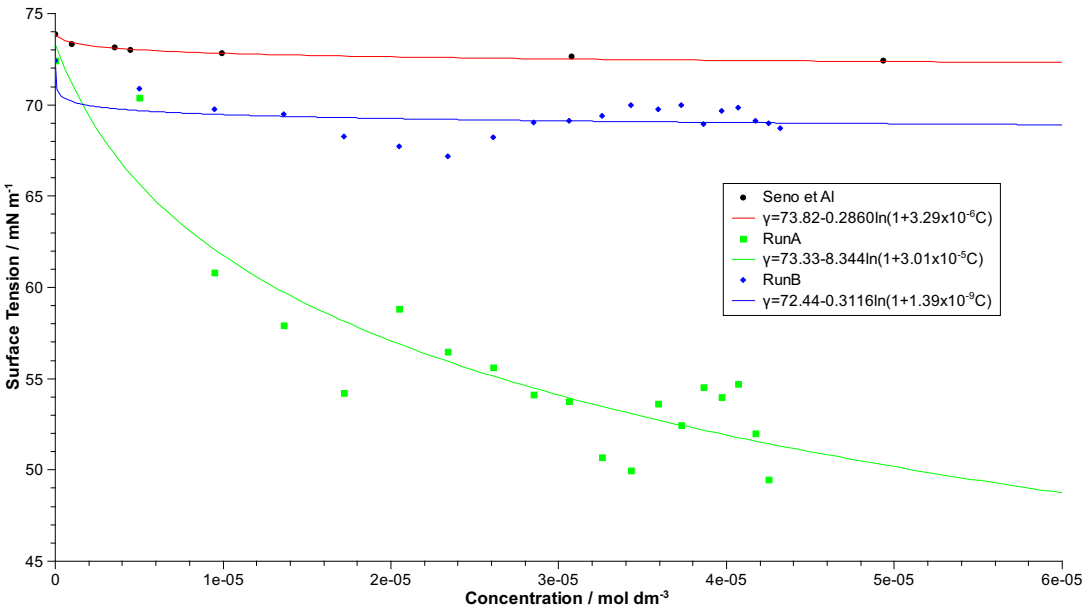


FIGURE 9.8: Surface Tension of Aqueous R6G solution.
Comparing local data to literature values. Data fitted to equation 2.36.

Polarisation dependence and concentration dependence measurements
Experiments were performed on a range of samples as outlined in chapter 7. Apparatus configuration was refined for an example of each concentration, and the most restrictive settings used for the range of samples covered by an experiment.

9.3 Results

9.3.1 Surface Tension

Results from data collected by various people at Southampton were collated and these proved to be inconclusive, figure 9.8. Measurements range from those agreeing closely with literature values to those with significant differences, even though all the Southampton data were collected with the same apparatus (a NIMA trough with Whilmey plate balance). When considered in isolation, some of the Southampton data present surface coverages implying molecular sizes of the same order of magnitude seen in the solid crystal structure, this is summarised in table 9.1. For the derivation of surface coverage later in this work, the values published by Seno *et al*[80], as presented in section 9.1.3, will be used.

The wide range of values measured suggests either that R6G does not behave in a predictable fashion, or that it is very sensitive to minor variations in experimental technique.

Dataset	γ_w (mN m^{-1})	$RT\Gamma_\infty$ (mN m^{-1})	k (mol^{-1} dm^3)	Γ_∞ (molecules m^{-2})	$\sqrt{\frac{1}{\Gamma_\infty}}$ (Å)
Literature	73.8	0.286	-3.29×10^6	7.02×10^{16}	37.74
Run A	73.3	8.34	3.01×10^5	2.05×10^{18}	6.99
Run B	72.4	0.312	1.39×10^9	7.65×10^{16}	36.16

TABLE 9.1: Surface Tension of Aqueous R6G solutions.
Comparing fitted co-efficients to literature values.

It is believed that R6G is highly surface active and moderately hydrophobic, hence it is likely that it will surface coat onto any solid that the solution contacts and indeed observations whilst working with R6G solutions show this as it is common to find apparatus becoming coated with a pink film. In the case of the surface tension experiment, this could occur on the paper Whilmey plates used with the tension balance. As these plates were subject to gravity as well as surface tension, deposition of the compound onto the plate could have changed the mass readings.

9.3.2 Ellipsometry

Work with Greef in the School of Engineering has produced the results presented in the 2007 paper[81]. The concentration and frequency dependence of the ellipsometry signal was studied. Analysis of the ellipsometric parameters and fitting against a three layer model of the surface layer has yielded information on the refractive index at the surface. The three layer model describes the surface as being a distinct layer bounded by air and a uniform bulk solution.

When left to minimise without restriction to the parameters, the model fitting tends towards the conclusion that the surface layer contains a high concentration of dye, often approaching 100 % dye. Initial fits also had a tendency towards unfeasibly high refractive index values. To control this, a bound was placed on the fit parameters, limiting the surface parameters to that of the solid dye previously measured. This limitation resulted in refractive index values close to that of the solid dye across the concentration range. This could be a consequence of the fitting procedure trying to force to higher refractive index values, but is consistent if it is accepted that the surface layer was nearly 100 % dye.

The refractive index was found to be constant within the precision of the experiment across the concentration range studied here, with the values presented in figure 9.9. These values were then fixed in the model which was used to probe the surface thickness and composition.

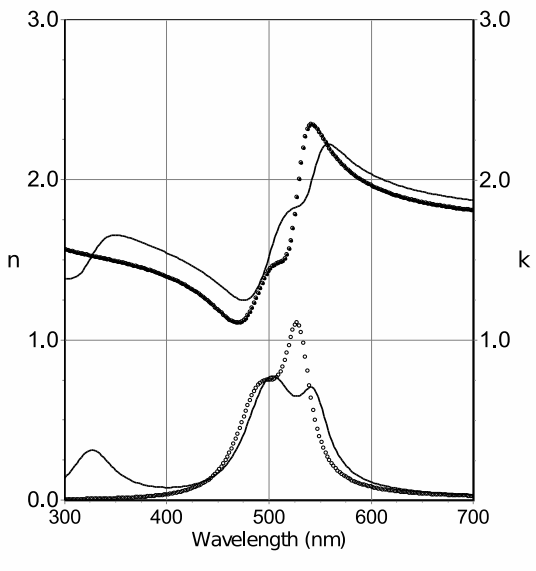
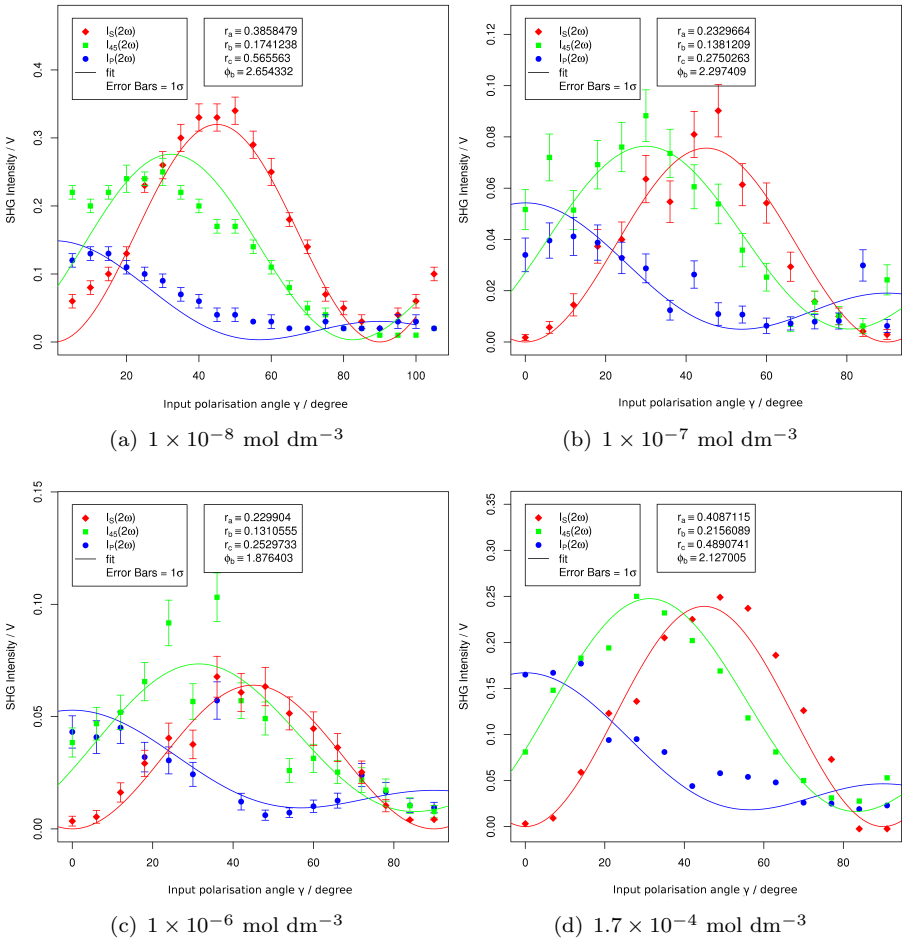


FIGURE 9.9: Refractive Index derived from ellipsometric measurements of R6G. *Taken from the 2007 paper[81]. Solid curves are for a solid R6G surface, dashed curves for a R6G and water surface. The lower pair of lines are the k values.*

9.3.3 SHG Polarisation results

Rhodamine 6G solutions were studied at various concentrations, with the aim of measuring the polarisation response. The results from these measurements were collated and compared, allowing for the production of a set of results containing modal behaviour for each concentration. For most of the concentration range studied, typical behaviour was easily found for each concentration, with atypical responses usually being caused by fault or failure in the apparatus leading to either no signal being measured or the measured signal being polluted with other (non-SHG) components. All the results presented as being typical have been measured a number of times on different days. A significant number (maybe double that presented) of other experiments monitoring polarisation were performed and not included, as they covered only a subset of the polarisation range needed for a reliable fit; these measurements were performed when testing the apparatus (for functionality and repeatability). Many show qualitative agreement with the typical data presented for the polarisation ranges they cover, hence the presented results should be considered with this repeatability in mind.

Figure 9.10 shows these results graphically. It will be noted that there are no data presented between 1×10^{-6} mol dm^{-3} and 1×10^{-4} mol dm^{-3}. In this region there was no clear modality in the data, although some repeatability was found. Figure 9.11 shows some examples of the data collected, it will be noted that there was generally a poor fit to the theoretical model for SHG polarisation, but that these data show the s-out curve to be the familiar shape, the 45^o-out curve to show a u shape between $0^o \leq \gamma \leq 90^o$ and the p-out curve to be poorly defined. The relative intensities of the curves change



(a) 1×10^{-8} mol dm^{-3} (b) 1×10^{-7} mol dm^{-3}

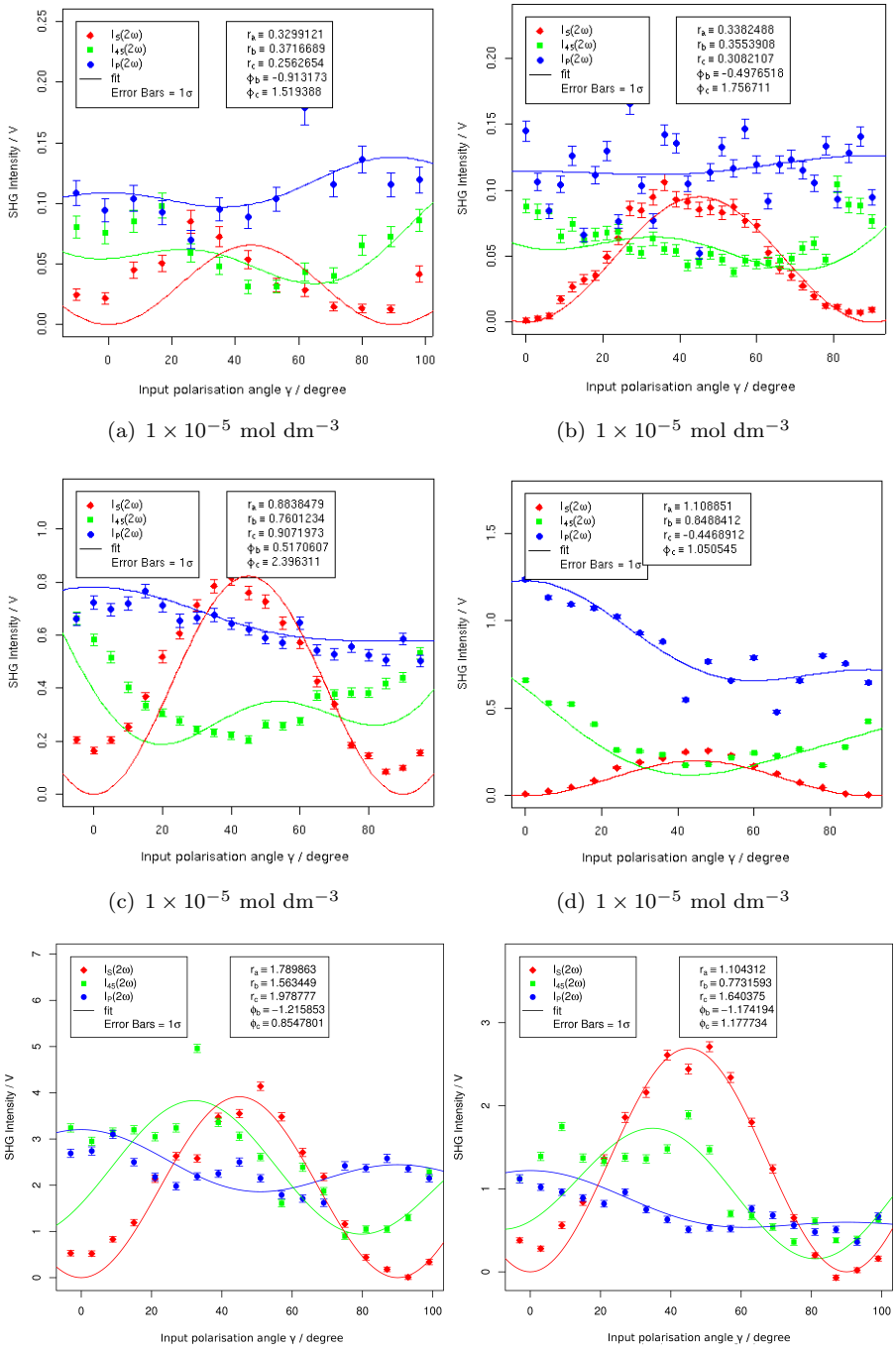
(c) 1×10^{-6} mol dm^{-3} (d) 1.7×10^{-4} mol dm^{-3}

FIGURE 9.10: Typical R6G polarisation scans at various bulk concentrations. Concentrations for which a typical behaviour has not been found have been excluded. Data fitted to equations 7.13 and 7.14. Error bars shown are standard error in the mean of the dataset represented by each point on the plot.

significantly between runs. At bulk concentration greater than 1×10^{-4} mol dm^{-3}, a return to repeatable, fitable, data is seen.

9.3.4 SHG Concentration Isotherm

An SHG concentration isotherm is built up by repeatedly recording measurements of a sample, and between each one exchanging some of the sample either for pure solvent (reducing the concentration) or stock solution (increasing the concentration). At the end of the experiment, the sample volume is checked, and the concentration at each step calculated. By inspection it should be obvious that the SHG signal will change with concentration of molecules at the surface. Assuming that the increase in surface coverage leads only to an increase in density of molecules, while their orientation stays the same, then one would expect a doubling in coverage to lead to a doubling of signal strength. Unfortunately, coverage cannot be measured directly, but it can be derived from the



(a) 1×10^{-5} mol dm^{-3}

(b) 1×10^{-5} mol dm^{-3}

(c) 1×10^{-5} mol dm^{-3}

(d) 1×10^{-5} mol dm^{-3}

(e) 4.5×10^{-5} mol dm^{-3}

(f) 4.5×10^{-5} mol dm^{-3}

FIGURE 9.11: Examples of R6G polarisation scans at various bulk concentrations for which a typical behaviour has not been found. Data fitted to equations 7.13 and 7.14. Error bars shown are standard error in the mean of the dataset represented by each point on the plot.

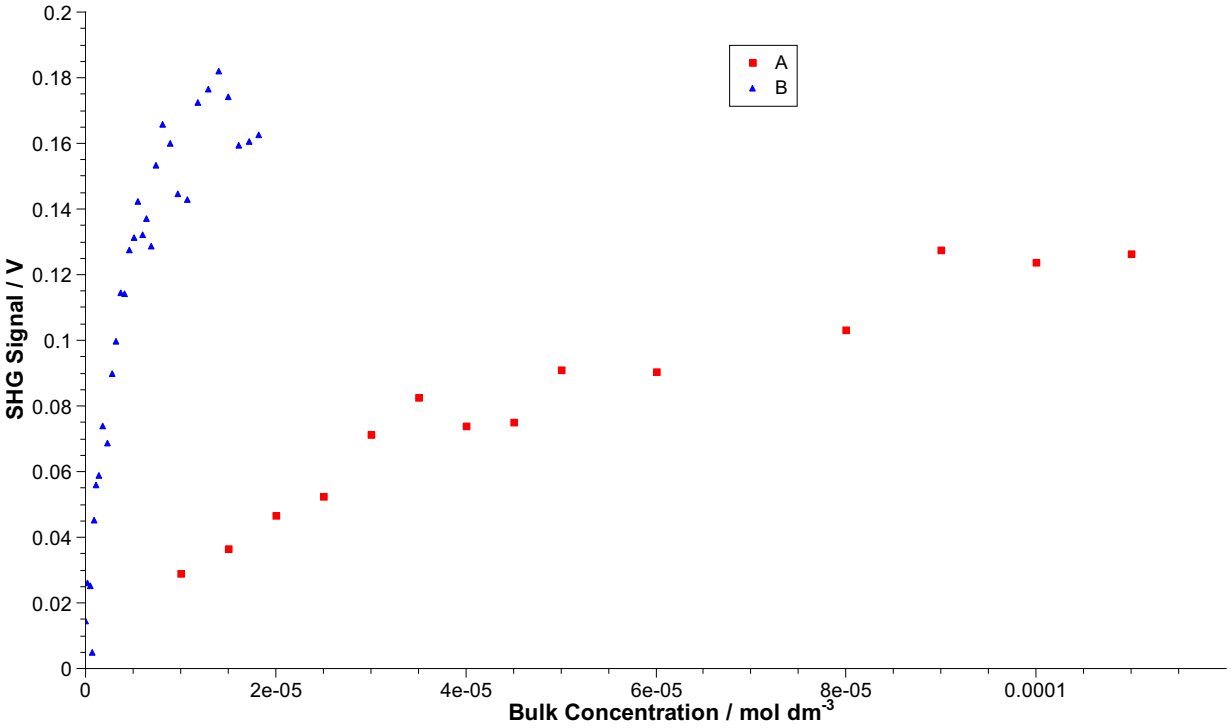


FIGURE 9.12: Raw data from a concentration dependent SHG experiment.
Showing the difference between runs that must be accounted for before comparison.

bulk concentration by surface tension measurements. Equation **??** shows that coverage is related to bulk concentration by a Langmuir relationship; hence for this simple case where it is assumed that SHG signal is directly proportional to coverage, it should be possible to fit a Langmuir to the SHG vs concentration data. If a poor or partial fit is found then this may indicate some deviation from the simple surface model, i.e. that there is some change occurring at the surface other than a simple increase in molecular coverage. Alternatively, the SHG data may be scaled against the coverage result from Surface Tension measurement, in which case deviation from the simple coverage/signal model is shown as deviation from linearity.

Experimental parameters were typically set to get a good compromise between signal and noise across the concentration range measured, hence may be different between isotherms. Some way of combining these separate data sets was then needed. Figure 9.12 shows examples of two separate concentration dependent SHG experiments. It is clear to see that there are some differences between them, however, this difference is in part due to the different experimental parameters used for the two runs. In order to obtain reliable SHG data from higher concentrations, it was necessary to run the apparatus under different conditions to those used for lower concentrations; there was a need to exclude the additional radiation from two photon fluorescence and to stay within

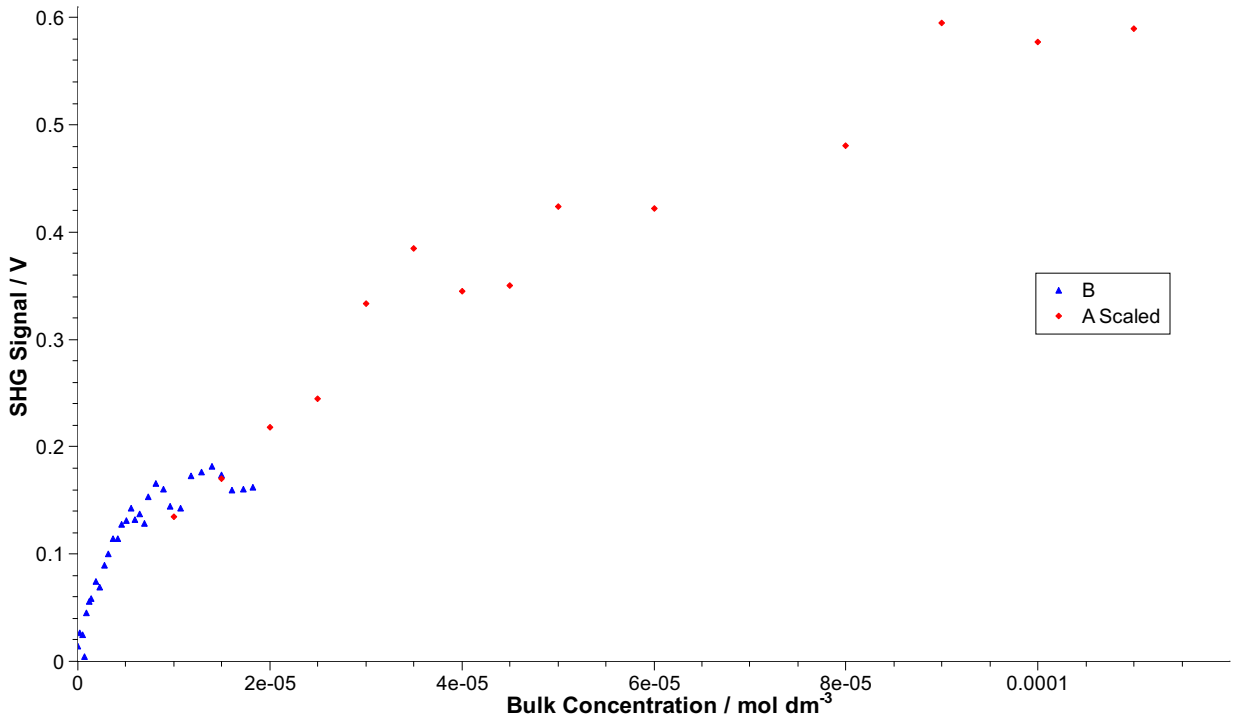


FIGURE 9.13: Scaled data from a concentration dependant SHG experiment.
Run A was scaled to to be numerically similar to B in the overlap region.

the capabilities of the detector system. Fortunately, these differences in conditions can be considered as simply changing the intensity of the signal, either at the input wavelength, the second harmonic wavelength or both. Equation 7.8 shows the link between the intensity of the input radiation I_ω and the intensity of the second harmonic $I_{2\omega}$. Attenuating the signal experimentally can be represented as a linear scaling of the (unit-less) signal terms, hence an attenuation multiplier can be inserted for both the input and output intensities. This equation can then be rearranged, and the two attenuation terms combined to form a single experiment attenuation term.

When comparing datasets, as in 9.12, an attenuation multiplier can be applied to SHG intensity in one dataset so that they overlap; hence it was possible to produce figure 9.13. In this case, the overlap region is relatively small; from initial inspection of this figure it would be very easy to conclude that there is some discontinuity in the overlap region, and that two Langmuir's, one based on dataset A and one of dataset B could be fitted. However, caution must be exercised to confirm that the scaling is in fact correct, as it is possible that at the high and low intensity regions of the experiment the detector system was not producing a linear response. Confidence in the correctness of this scaling can be enhanced by the addition of other overlapping datasets.

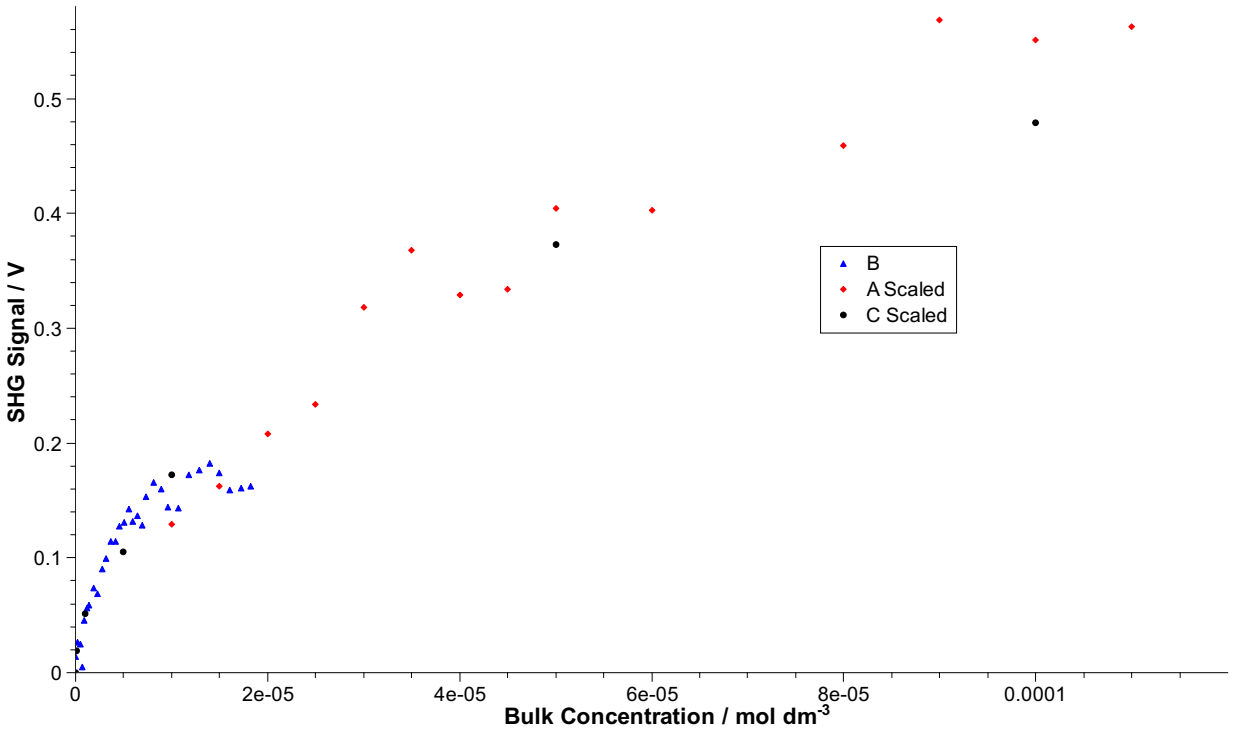


FIGURE 9.14: Combining many concentration dependant SHG experiments improves confidence in the data.

Dataset C (scaled to match) is added in figure 9.14. This indicated that the previous scaling on A was reasonably accurate, although was tweaked slightly in the production of figure 9.14. By inspection alone, the discontinuity in surface tension against concentration now appears quite probable, as was suggested by Danos[66], and by the fits performed on the literature surface tension data.

9.4 Discussion

9.4.1 Surface Tension

As the surface tension results collected in Southampton were so inconclusive, it was decided to use only the previously published data and consider them more closely (figure 9.15). In this figure, subsets of the literature data are considered and fitted against the Langmuir equation as derived in section 2.2, the conclusion of which (equation 2.36) is reproduced here for reference.

$$\gamma = \gamma_w + RT\Gamma_\infty ln(1 + kC) \tag{9.1}$$

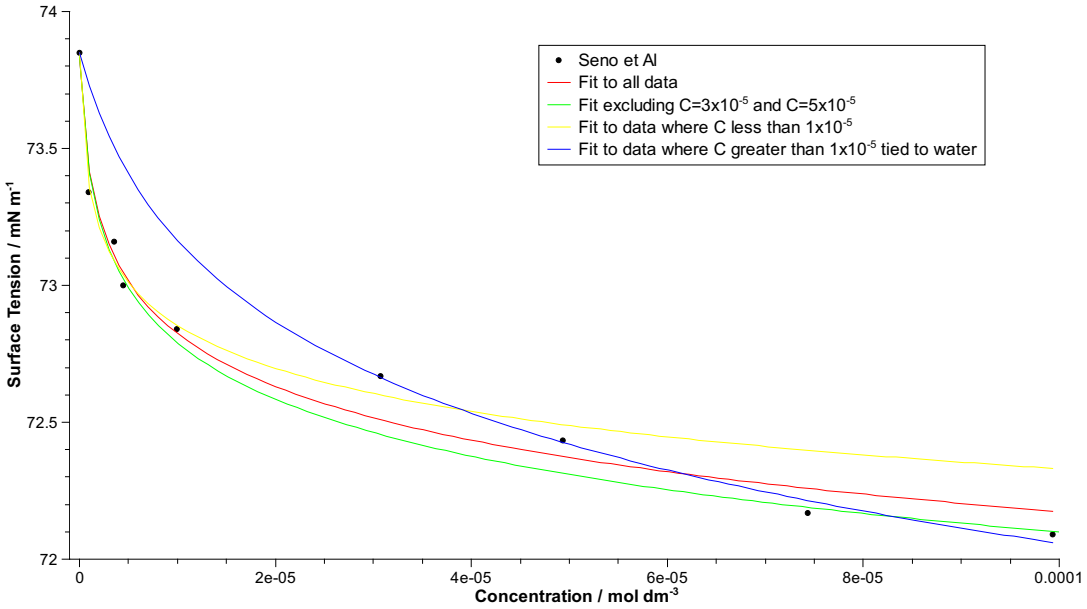


FIGURE 9.15: Surface Tension of Aqueous R6G solution.
Fitting the literature data to Langmuir isotherms (equation 2.36).

The fit presented as background material (red line) shows that there is some variation from the Langmuir fit at higher concentrations. At a first glance, it might be reasonable to think that the two points at 2×10^{-5} mol dm^{-3} and 5×10^{-5} mol dm^{-3} are the outliers from the Langmuir that is being fitted. Hence the second fit (green) is based on the data with these two points removed. Visually this appears valid, however care must be taken not to pick a fit that appears effective, and justifies our hypothesis without exploring the data completely. It is possible that a change occurs in the system's behaviour at around 1×10^{-5} mol dm^{-3} and that all data beyond this point should be considered separately. The third (yellow) fit considers just the points below 1×10^{-5} mol dm^{-3}, suggesting in fact that the 2×10^{-5} mol dm^{-3} and 5×10^{-5} mol dm^{-3} data may be more correctly aligned with the low concentration than the two higher concentration points.

Unfortunately, the dataset extracted from the literature does not contain many points, so the conclusions drawn here are fairly speculative. By considering the low concentration data independently from the high concentration data, on the basis that two independent systems being displayed and that the transition between them occurs around 1×10^{-5} mol dm^{-3} solutions as appears to be the case, then attempting to fit the higher concentration data would be sensible. As this is ultimately a water solution then a "water only" point should be included to limit the fit to a reasonable intercept. This final fit (blue) fits the 4 high concentration points well, which conveniently fits with the hypothesis that there is a change occurs in the sample's surface in this concentration region.

Dataset	γ_w (mN m^{-1})	$RT\Gamma_\infty$ (mN m^{-1})	k (mol^{-1} dm^3)	Γ_∞ (molecules m^{-2})	$\sqrt{\frac{1}{\Gamma_\infty}}$ (Å)
All data	73.8	0.286	-3.29×10^6	7.02×10^{16}	37.72
Excluding 2×10^{-5} mol dm^{-3} and 5×10^{-5} mol dm^{-3}	73.8	0.304	3.03×10^6	7.46×10^{16}	36.62
Ignoring all data above 1×10^{-5} mol dm^{-3}	73.8	0.229	7.61×10^6	5.62×10^{16}	42.19
Data above 1×10^{-5} mol dm^{-3} and water	73.9	0.554	2.45×10^5	1.36×10^{17}	27.13

TABLE 9.2: Surface Tension of Aqueous R6G solutions.
Comparing fitted co-efficient values

If the numerical data that can be derived from the fits are considered, the summary in table 9.2 can be produced. This includes the derived "molecular size", assuming that the molecule is approximately square. The X-Ray diffraction data for solid R6G gives a molecule which is approximately square in the plane of the xanthene core with sides 10 Å long, hence surface tension data should give molecular sizes in this order of magnitude. However, it is found that data indicates molecules about four times larger than this, much more than can be accounted for by intermolecular forces (e.g. hydrogen bonding, typically 1.5 Å to 2 Å), or hydration spheres (the water molecule can be considered to be add at most 3 Å, 1 Åfor the OH bond, and 2 Åhydrogen bonding the water to the molecule). It was concluded that, assuming the surface tension data to be accurate, some other molecular re-arrangement was occurring at the surface. Given the low number of points in these data, it was impossible to draw definite conclusions, but the fits presented (figures 9.15 and 9.2) suggest that either the data at 3×10^{-5} mol dm^{-3} and 5×10^{-5} mol dm^{-3} are outliers to the rest of the data (green fit), or that there is some change in the behaviour of the data between 1×10^{-5} mol dm^{-3} and 6×10^{-5} mol dm^{-3}, and that separate Langmuir isotherms should be fitted for the two regions (yellow and blue fits).

9.4.2 SHG Concentration Isotherm

Concentration dependant SHG data were collected over a number of experiments and combined as described in section 9.3.4. Langmuir isotherms were then fitted to the data,

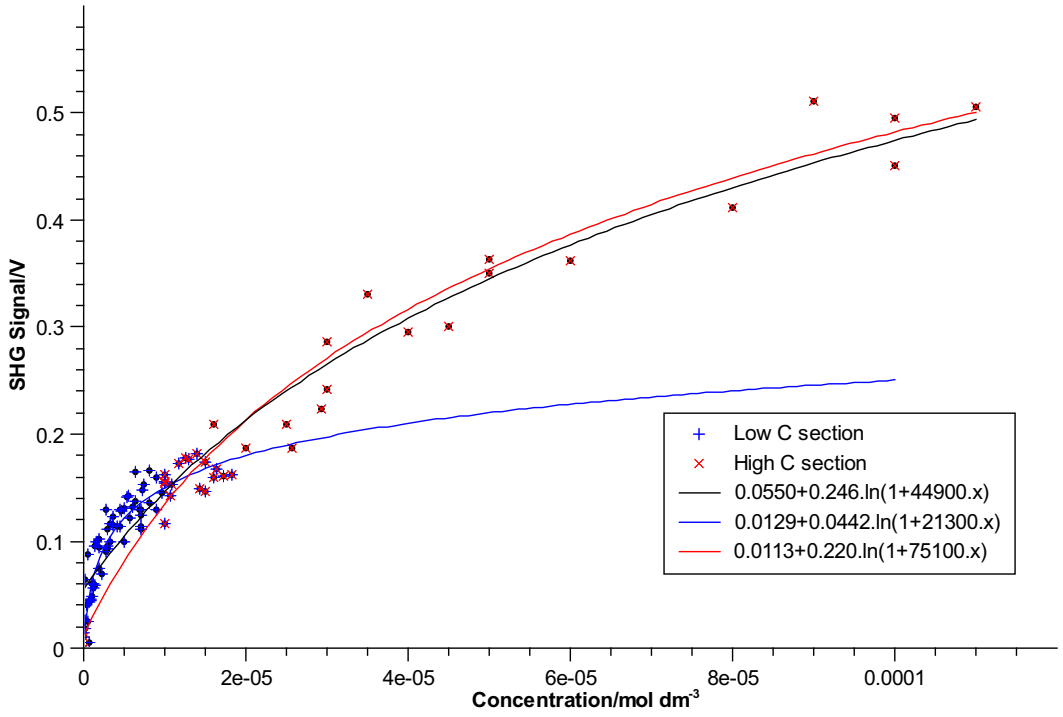


FIGURE 9.16: Analysing the concentration dependant experiment.
Fitting Langmuir isotherms (equation 2.36) to the data suggests that it falls into two sections.

figure 9.16, to aid interpretation. An initial fit to all the data (black curve) gives a model that agrees well with the data at higher concentrations, but doesn't intercept correctly and fails to model the lower concentration data well. By separating the data into two sets with an overlap, much more representative Langmuir models were fitted; this produced a high concentration curve similar to that fitted to the complete data set(red), and a very different good fit to the lower concentration data (blue). The low concentration fit suggests that system reaches an initial coverage saturation around 1×10^{-5} mol dm^{-3} and 2×10^{-5} mol dm^{-3} before overcoming some energy threshold that allows for a molecular rearrangement to the system modelled by the "high concentration" (red) curve.

This change in surface behaviour agrees with the tentative conclusions from the surface tension data, which also indicated a possible change in surface behaviour between 1×10^{-5} mol dm^{-3} and 6×10^{-5} mol dm^{-3}. When viewed in light of the SHG isotherm data, the second model proposed by surface tension data (that the data should be modelled in two sections) becomes more viable but that there is some disagreement about where the transition occurs.

Figure	Bulk Concentration (mol dm^{-3})	A	B	C	$\frac{B}{A}$	$\frac{C}{A}$
9.10(a)	1.00×10^{-8}	0.386	0.174+2.654i	0.566	0.451+6.879i	1.466
	1.00×10^{-8}	0.624	0.282+1.448i	0.531	0.452+2.321i	0.852
	1.00×10^{-7}	0.481	0.232+1.052i	0.415	0.482+2.189i	0.864
	1.00×10^{-7}	0.199	0.177+1.833i	0.259	0.889+9.230i	1.302
9.10(b)	1.00×10^{-7}	0.233	0.138+2.297i	0.275	0.593+9.861i	1.180
	1.00×10^{-7}	0.203	0.137+2.043i	0.246	0.675+10.044i	1.209
9.10(c)	1.00×10^{-6}	0.230	0.131+1.881i	0.250	0.565+8.174i	1.087
	1.00×10^{-6}	0.160	0.064+1.268i	0.089	0.375+7.938i	0.563
9.11(a)	1.00×10^{-5}	0.319	0.357	0.256+2.043i	1.120	0.802+6.397i
	1.00×10^{-5}	0.829	0.593+1.078i	0.397+2.357i	0.716+1.299i	0.479+2.841i
	1.00×10^{-5}	0.948	0.890+1.386i	0.556+4.191i	0.939+1.463i	0.587+4.422i
9.11(d)	1.00×10^{-5}	1.101	0.827+0.959i	0.479+2.779i	0.752+0.871i	0.435+2.526i
9.11(b)	1.00×10^{-5}	0.338	0.346	0.308+2.0128i	1.025	0.913+5.957i
9.11(c)	1.00×10^{-5}	0.867	0.755	0.908+2.1579i	0.871	1.047+2.487i
	3.00×10^{-5}	0.383	0.472+0.792i	0.326+0.784i	1.233+2.071i	0.853+2.049i
9.11(e)	4.50×10^{-5}	1.790	1.563-1.216i	1.979+0.855i	0.873-0.679i	1.105+0.478i
9.11(f)	4.50×10^{-5}	1.039	0.874+2.225i	1.583	0.841+2.142i	1.524
	1.70×10^{-4}	0.355	0.176+2.294i	0.490	0.494+6.455i	1.379
9.10(d)	1.70×10^{-4}	0.409	0.216+2.127i	0.489	0.528+5.204i	1.197

TABLE 9.3: Model Co-efficients for R6G polarisation SHG data.

9.4.3 SHG Polarisation Dependence measurements

Table 9.3 contains the fit co-efficients for the SHG polarisation model presented in section 7.1.5. The data presented here are typical of those for each concentration, and were manually selected from the complete set of data collected. Fits were performed on the data, limiting combinations of the B and C co-efficients from having imaginary components, the resultant fit parameters were then examined. Whilst the lowest residual sum of squares value for the fit was usually found when both B and C are allowed imaginary components, typically, the imaginary part of one of these had only a small effect on the fit hence, where this is the case, the values presented represent that coefficient (B or C) being forced real.

It should be noted that the imaginary part of the C coefficient only becomes significant in the 1×10^{-5} mol dm^{-3} data. As mentioned when the data were first presented there is much uncertainty in the data at this concentration as the fit is generally poor and it is unclear if the lack of fit is a problem with the model, the sample or the data

collection. In the case of R6G solutions, inspection of the raw data plot suggests that the C imaginary component apparently becoming more significant is actually a display of the fit's accuracy becoming poorer, and requiring more degrees of freedom to attempt to get close to the experimental data. It would be more appropriate to consider a different model for these data, however the poor repeatability of these data should first be addressed.

Given the care taken with these experiments to provide reproducible conditions (including some of the experiments being run sequentially on the same day) and the irreproducibility still seen, it is reasonable to suggest that there is some inherent instability in the surface. Studies of bulk aqueous R6G solutions have shown there to be a transition in bulk behaviour at around this concentration (from monomer to dimer aggregates), hence it is likely that changes are also occurring at the surface. It is quite possible that, within the time frame of this experiment, the surface under the laser spot changed, possibly multiple times. When the fit coefficients are compared numerically, it can be seen that across all the concentrations studied there is significant variation between different measurements on samples of the same makeup. In a perfect experiment, one would expect these differences to be negligible and would hope to get the same result for two repeats of the same measurement. For data, like this, where steps have been taken to remove variation between the measurements (both during experiment and in the data work-up), the expectation is to find very similar results for repeated measurements. Where this is not the case, there are two possible conclusions either that the experiment is unreliable or that the sample is inherently unstable.

In order to prove the reliability of the experiment, study of a much simpler system without multiple component interactions can be undertaken. As R6G was studied in an aqueous environment, a study of pure water samples was performed and the results presented in the next chapter.

Chapter 10

Water

10.1 Background

Water is an abundant part of human existence and everyday life; over 71 % of the earth's surface is covered with water[82]. Without it, many of the biological and mechanical processes that we take for granted on a daily basis would cease to function. Therefore, an understanding of its behaviour underlies the study of many complex processes.

In this work, the air/water interface is considered. Chemistry at aqueous interfaces underpins many biological systems, the oil/water interface being a common model for the lipid interface. To aid understanding of these systems, a thorough understanding of aqueous interfaces is desirable. Whilst most complex systems include a range of compounds in competition, the study of most aqueous systems references back to pure water to provide a baseline or intercept, especially when looking at concentration-dependent behaviour.

When studying aqueous processes, it is often desirable to examine low concentration systems in which the water component makes an important contribution to the overall signal. In this case, knowledge of how water behaves will permit consideration of how water-like the sample is.

Water is a useful standard within the SHG experiment, especially when considering aqueous samples. Assuming that there are no contaminants, a pure water sample gives a repeatable, although small, SHG response. Being able to measure this response provides a good indication of the sensitivity of the apparatus and its tolerance to noise (both optical and electrical). The SHG response of a particular water sample can also give a good indication of the presence of surface active contaminants that may interfere when recording the response of a sample. This is particularly important when dealing with very low concentrations of sample (when the amount of contaminant becomes significant with respect to the amount of sample).

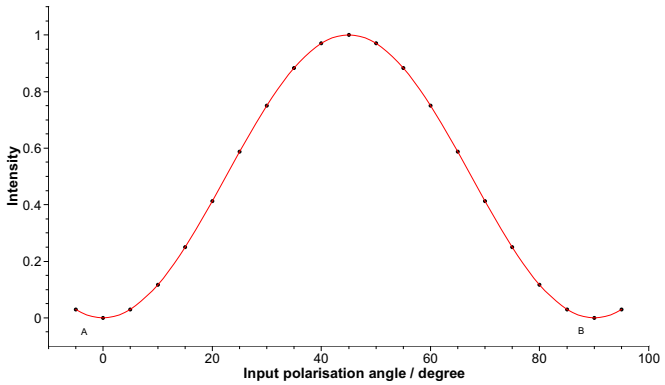


FIGURE 10.1: S_{out} SHG polarisation dependence for achiral samples (equation 7.14).

In section 7.1.5 the following equation is presented for the polarisation dependence of the intensity of S_{out} second harmonic for achiral samples.

$$I_s^{(2\omega)} \propto |C \sin 2\gamma|^2 I_\omega^2 \tag{10.1}$$

It can be seen from this that at an input polarisation of 0^o or 90^o the second harmonic intensity will be zero. This will continue to be the case when there are achiral contaminants present.

If it is expected that there will be chiral molecules at the surface, then the analysis presented by Hecht and Barron should be followed[83]. It was presented that for chiral molecules the polarisation dependence of the S_{out} second harmonic becomes:

$$I_s^{(2\omega)} \propto \left|C \sin 2\gamma + D \cos^2 \gamma\right|^2 I_\omega^2 \tag{10.2}$$

In this case it can be seen that for γ=90^o, intensity will be zero, but for γ=0^o, intensity will be non-zero.

Based on this knowledge some simple diagnostics for the S_{out}-dependence alone can be defined, allowing for some incorrect conditions to be detected without further investigation. If we consider the plot in figure 10.1 and evaluate the data at points A and B, then the inferences in table 10.1 can be defined. It should be noted that at higher powers there is also the possibility of finding higher than second order optical effects. At high powers, there is also the danger of damaging the samples or optics, hence the experiment is usually optimised to as low a power as is considered to give a good signal to noise ratio for the sample.

Previous work by Frey *et al*[71] investigated the link between temperature and second harmonic signal at the air/water interface. Measurements of the second harmonic generated at the water/air interface for a range of sample temperatures were made and it was

I(A)	I(B)	Inference
≈ 0	≈ 0	Expected SHG response for achiral components at the surface.
$\neq 0$	≈ 0	Chiral contaminant, and/or higher order effects.
$\neq 0$	$\neq 0$	Offset (e.g. electrical or optical), and/or higher order effects.

TABLE 10.1: Achiral S_{out} diagnostics table.

found that second harmonic generation decreases with elevated temperature. A more thorough investigation of the polarisation dependence of the signal was made at room temperature. It was found that despite the experiment being non-trivial to perform, reproducible data could be collected and that it was in agreement with the theoretical model, producing a good fit for real values of the A, B and C parameters.

10.2 Experimental

Water samples from three sources were studied using the range of experimental techniques discussed in chapter 7. Some data were collected before and during the work on R6G, and some as a result of the inconclusive nature of some of the R6G results. Sources 1 and 2 were from the same research group, 1 being an earlier system that was replaced part way through this work, source 2 being their new system. Their work focuses on micro-crystalline structures being used as catalysts, and the water is considered to be very low in organic contaminants. The third source is from a group researching the optical properties of samples deposited onto solid surfaces, and the water is required to be very low in contaminants as they work with low concentration samples.

Specifically, the diagnostic techniques of monochromator scanning and power dependence were used to check the viability of the experiment, thus allowing polarisation scans to be performed with confidence. Additionally, subsets of the polarisation scan were repeated to test the repeatability, and further improve confidence. Figures 10.2, 10.3, 10.4 and 10.5 show representative spectra recorded for each of these tests. The error bars in these figures are the standard error of the raw data set of which that data point is the mean. The results from these experiments were compared, firstly by qualitative inspection for variations in the curve shapes with sample, and secondly by comparing the numeric values obtained as fit parameters when fitting against the theoretical models (as presented in section 7.1).

The monochromator scan (figure 10.2) shows an SHG peak that is large compared to the signal floor. This is the near ideal case for performing SHG experiments, as the monochromator slits width can be increased, allowing all the second harmonic radiation to be captured without the worry of stray radiation producing an intensity offset.

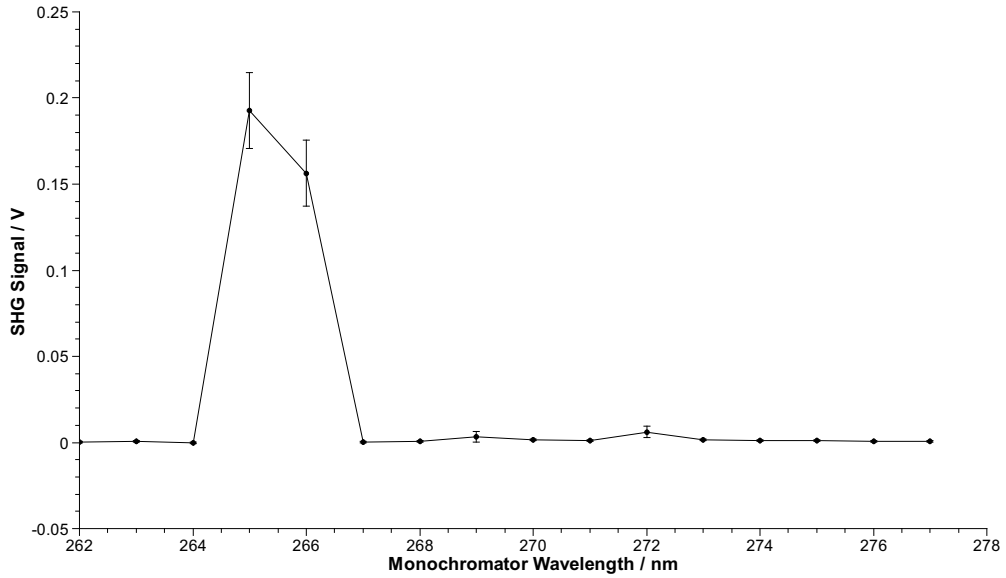


FIGURE 10.2: Sample monochromator scan for water. Showing a strong peak at 266 nm, and little signal at near wavelengths. Error bars shown are standard error in the mean of the dataset represented by each point on the plot.

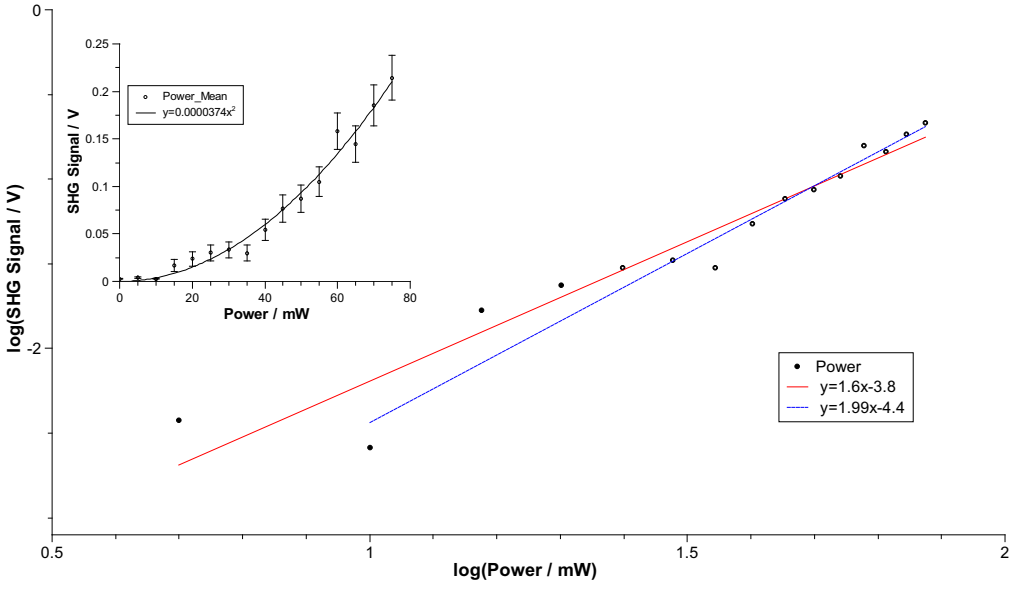


FIGURE 10.3: Sample power dependence of water SHG. Showing the $SHG \propto Intensity^2$ as predicted by the theoretical analysis of the second harmonic behaviour. Error bars shown are standard error in the mean of the dataset represented by each point on the plot.

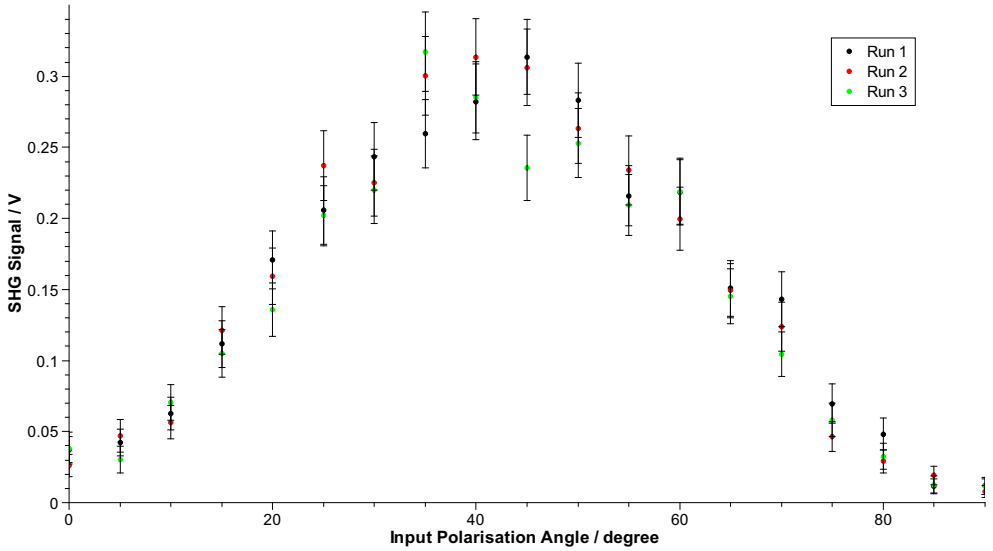


FIGURE 10.4: An example of a "repeated polarisation" experiment. This experiment holds the output polarisation static over a number of repeats. This experiment shows the data collection is reproducible within an acceptable error. Error bars shown are standard error in the mean of the dataset represented by each point on the plot.

Figure 10.3 shows the power dependence of the second harmonic generated from a water surface; again this is as was theoretically predicted, fitting to a quadratic equation within the experimental uncertainty of the data (as shown by the error bars). This is confirmed by a plot of Log(Intensity) against Log(power) which shows a gradient of approximately 2 for data with an input power greater than 25 mW. When including lower power data in the linear fit, the gradient drops to 1.6, indicating that at lower powers either the detector is picking up additional light or that the power meter is reaching its lower sensitivity limit.

The repeated polarisation experiment (figure 10.4) is a special case of the polarisation scan, in which the output polariser is not moved between scans of the input polarisation. For a stable, repeatable, system one would expect to record the same values for each scan, as shown. As this experiment was performed over the same timescale as a polarisation scanning experiment, its results also show the system to be stable over the time taken for such a scan.

Figure 10.5 shows an example of the polarisation response measured for a water sample. At the wavelength studied (532 nm source, 266 nm second harmonic) there should be no significant resonance enhancement to the second harmonic response of a water sample, hence the A, B and C parameters can be assumed to be real[71] as shown in this example. If a poor fit when A, B and C are assumed real improves when B and C are allowed an imaginary component suggests that there is some contaminant at the surface. It should be remembered though that water isn't the only compound for which all the fit

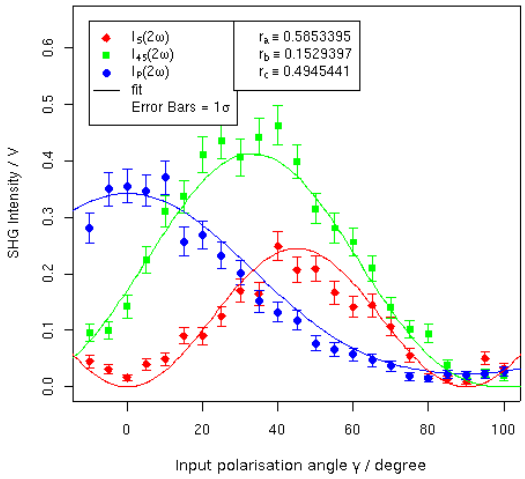


FIGURE 10.5: Sample fitted polarisation plot for a recent water sample (Water Source 2). Data fitted to equations 7.13 and 7.14. Error bars shown are standard error in the mean of the dataset represented by each point on the plot.

parameters can be real, hence the absence of imaginary components alone doesn't prove the sample to be pure water.

10.3 Results

The diagnostic measurements presented in the previous section show that, for water samples, the apparatus is producing data with a good signal to noise ratio for moderate input powers. It is also shown to be reproducible within the time period of a polarisation dependence data collection. Measurements were made for a number of samples, these results were then fitted to the theoretical model. For each experiment, the data were fitted forcing all the parameters to be real. The model was then relaxed, firstly allowing C to be complex, then B to be complex and finally allowing both B and C to be complex; when presenting these results, the fit model used is identified by listing the real parameters in the fit. An analysis of the variance between these results was performed (ANOVA F-test) to quantify the significance of making the parameters complex (i.e. between ABC real, and each of AB real, AC real and A real). For F-test showing less than 95 % confidence that allowing complex parameters is significant, the fit was ignored. Where more than one model showed greater than 95 % confidence, the most significant model (i.e. that with the highest confidence) was chosen. Table 10.2 shows the results of this analysis using water source 1 as an example, and table 10.3 presents the most significant model fits across all the water sources.

Experiment	Fit Type	A	B	C	Residual Sum of Squares	ANOVA	
						F Test	Pr(>F)
1	ABC	0.296	-0.0365	0.322	0.0142		
	AB	0.300	-0.0153	$0.325 - 0.615i$	0.0138	1.40	0.242
	AC	0.296	$0.102 + 2.00i$	0.322	0.0136	2.50	0.119
	A	0.298	$0.0959 + 1.538i$	$0.328 - 0.581i$	0.0129	2.95	0.060
2	ABC	0.164	0.00867	0.168	0.000354		
	AB	NoFit	NoFit	NoFit			
	AC	0.165	$0.0480 + 1.59i$	0.168	0.000338	1.35	0.254
	A	0.166	$0.0510 + 1.70i$	$0.168 + 0.189i$	0.000334	0.843	0.441
3	ABC	0.692	0.153	0.338	0.0326		
	AB	0.693	0.154	$0.341 + 0.246i$	0.0325	0.112	0.740
	AC	0.697	$0.204 + 1.01i$	0.0.344	0.0300	4.87	0.0314
	A	0.703	$0.214 + 1.19i$	$0.351 + 0.525i$	0.0267	6.06	0.00418
4	ABC	0.374	0.185	0.147	0.0205		
	AB	0.389	0.204	$0.259 + 1.68i$	0.00615	81.90	1.08×10^{-10}
	AC	0.380	$0.214 + 1.04i$	0.166	0.0198	1.37	0.249
	A	0.392	$0.203 + 0.349i$	$0.258 + 1.78i$	0.00592	42.0	6.47×10^{-10}

TABLE 10.2: ABC values for water SHG polarisation dependence for source 1. *Comparing the effect of forcing different fit parameters real.*

On inspection of the data, the first impression is that source 2 mostly produces data that fits the model well with only real fit parameters, and that both sources 1 and 3 show some variability. Sources 1 and 3 sometimes produce good, real-only (ABC), fits but also produce results for which allowing B and/or C to become complex results in a significant improvement of the fit. When considering the significance of the fit type to the anomalous source 2 experiment, the significance of allowing C to be complex had a confidence level of approximately 99.5 %. Given that, for all the other source 2 experiments, allowing B and C to be complex is insignificant it is likely that a further effect is occurring, such as external contamination in the lab, and that this data point can be discarded from this discussion.

Given the high variability of the data above, the first conclusion to draw is that SHG is highly sensitive. The water sources are all classed as "high purity" systems, all have high resistivity readings ($\geq$ 18.0 MΩ cm), and all have equivalent surface tensions (within the sensitivity of the apparatus), hence as it is assumed that the cleaning protocol in use for the SHG experiment successfully removed all contamination from the sample holder and sample handling equipment, the variability in results can only be caused by variation in the sample and that the SHG apparatus is sensitive enough to detect variations when

Water Source	Fit Type	A	B	C	$\frac{B}{A}$	$\frac{C}{A}$
1	ABC	0.296	-0.0364	0.322	-0.123	1.09
1	ABC	0.164	0.00867	0.168	0.0530	1.02
1	A	0.703	$0.214 + 1.20i$	$0.351 + 0.525i$	$0.304 + 1.70i$	$0.499 + 0.746i$
1	AB	0.389	0.204	$0.259 + 1.68i$	0.524	$0.664 + 4.31i$
2	ABC	1.21	0.297	1.21	0.246	1.00
2	AB	0.901	0.0979	$0.901 + 0.678i$	0.109	$1.00 + 0.752i$
2	ABC	0.470	0.151	0.246	0.321	0.523
2	ABC	0.630	0.182	0.298	0.289	0.473
2	ABC	0.585	0.153	0.495	0.261	0.845
2	ABC	0.344	0.033	0.551	0.095	1.61
3	AC	0.348	$0.127 + 1.51i$	0.301	$0.364 + 4.33i$	0.866
3	ABC	0.385	0.0551	0.297	0.143	0.771
3	ABC	0.425	0.0874	0.348	0.206	0.821
3	A	0.167	$0.0556 - 1.63i$	$0.192 + 0.957i$	$0.333 - 9.76i$	$1.15 + 5.73i$
3	AB	0.526	0.147	$0.400 + 0.945i$	0.280	$0.760 + 1.798i$
3	AC	0.613	$0.176 + 1.30i$	0.620	$0.287 + 2.11i$	1.01

TABLE 10.3: ABC values for water SHG polarisation dependence.
Collating the water source data to allow comparison of the various sources.

the other tests do not. Had all the water sources shown the same level of variability, then it would have been possible to conclude that the variability was being introduced by the experiment. The possibility that the variability is in fact a compound effect should not however be discounted, with some of the variability being due to the sample and some due to the instrument.

Table 10.4 presents the mean and standard error of the three sources presented, and also the values published by Frey in 2001[71] for comparison. It should be noted that the standard error published with the Frey data was from the fit goodness, not the agreement of a number of fits. However, the size of the standard error found in the data from sources 1, 2 and 3 is of concern, by being of the same order of magnitude as the mean values it shows that very little confidence can be held in these data; the level of agreement is so low that it is impossible to generate any conclusions about what the actual ratio values should be. The cause of this disagreement should be considered so that improvements to the experiment can be designed in order to improve the confidence in future data sets.

Water Source	Value	$\bar{x}$	σ Real Part	σ Imaginary Part	$\lvert\sigma\rvert$
1	$\frac{B}{A}$	$0.190 + 0.425i$	0.246	0.736	0.776
	$\frac{C}{A}$	$0.818 + 1.26i$	0.245	1.78	1.80
2	$\frac{B}{A}$	0.242	0.0780		0.0780
	$\frac{C}{A}$	0.890	0.410		0.410
3	$\frac{B}{A}$	$0.269 - 0.553i$	0.0746	4.41	4.41
	$\frac{C}{A}$	$0.896 + 1.25i$	0.140	2.11	2.11
Frey 2001	$\frac{B}{A}$	0.266	0.01		
	$\frac{C}{A}$	0.722	0.007		

TABLE 10.4: Water SHG polarisation fit ratios.
σ values for the Frey data are based on variance in the fit, not on a set of fit parameters.

Looking back at the R6G data, it can be seen that for moderate concentration systems the experiment can produce data which has a reasonable level of agreement. This suggests that the cause of these variations is coming from within the sample or sample handling rather than from the instrument, or that the signal from water samples is so low that it is beyond the sensitivity of the instrument. Instrument sensitivity can be quickly discounted by the diagnostics presented for water, all of which had been repeated on a number of occasions. It was shown that it is possible to detect a strong second harmonic peak on a monochromator scan, and to show quadratic behaviour when testing the dependence of second harmonic generation on input power. Had there been a signal sensitivity issue, then both of these tests would have shown noise and it may not have been possible to detect the expected behaviour. This suggests either contamination or variation in the water sources, or variation in the cleanliness of the sample handling. The R6G data show that SHG is very sensitive to low concentration samples, collection of repeatable data for 1×10^{-7} mol dm^{-3} R6G solutions was performed as routine, and there was some confidence in the data collected for 1×10^{-8} mol dm^{-3} R6G solutions. Hence, when attempting to collect data on a pure solvent it should be obvious that the amount of contaminant needed to distort this data is minuscule, after all the signal from 1×10^{-7} mol dm^{-3} R6G is enough to override an equivalent contaminant signal. Given the care taken when performing the experiments, it is difficult to comprehend how this can be improved without the availability of "clean-room" facilities. If this type of contamination is the cause of these problems, then further repetition and averaging over a much larger dataset may help, although it may just confirm the unpredictability.

Contamination or variation of the source is harder to quantify, as the sources were not under the direct control of the author. It is likely that some degradation in purity of the source may occur as the filters and UV lamp in the purifier age and that, given a large

Water Source	Residual Sum Of Squares	A	B	C	$\frac{B}{A}$	$\frac{C}{A}$
1	0.0142	0.296	-0.0364	0.322	-0.123	1.09
1	0.000354	0.164	0.00867	0.168	0.0530	1.02
2	8.41	1.21	0.297	1.21	0.246	1.00
2	0.0360	0.470	0.151	0.246	0.321	0.523
2	0.118	0.630	0.182	0.298	0.289	0.473
2	0.0462	0.585	0.153	0.495	0.261	0.845
2	0.0819	0.344	0.033	0.551	0.095	1.61
3	0.00576	0.385	0.0551	0.297	0.143	0.771
3	0.00790	0.425	0.0874	0.348	0.206	0.821
3	0.00131	0.161	-0.0499	0.18584	-0.309	1.15
3	0.0256	0.610	0.0837	0.616	0.137	1.01

TABLE 10.5: ABC values for water SHG polarisation dependence, A,B and C assumed to be real.
Collating the water source data to allow comparison of the various sources.

enough sample set plotted against time, one may see an increase in the variability of results followed by sudden drop in variability corresponding to the filters being changed.

As has already been stated qualitatively, the presence of different fit assumptions producing best fits for a given source is good indication that there is variation within that source, as it suggests the presence of resonance enhancing contaminants. Even when the ANOVA results are reconsidered with a 99 % significance test, some of the results still show complex parameters as significant. If these are then discarded (based on the assumption that they contain a contaminant) table 10.3 reduces to table 10.5. The sample size for the water sources presented therein is much reduced, hence combination of all the data was used for some of the following analyses. Considering that there is still some doubt that all the samples were pure water, the presence of outliers is expected. A box-plot of the combined ratios (figure 10.6) makes two outliers quite obvious. By removing these the averages presented in table 10.6 are obtained. Source 1 is excluded from this table as the mean of two values with so much variance is quite meaningless.

Given the uncertainty in the mean ratios calculated, it is unclear if there is any difference in the sources. When considering the combined dataset, there is some agreement with published values although the error shown makes this quite speculative. It should be remembered that the standard error quoted by Frey is significantly lower than those calculated for these data, as it is based on the fit variance rather than the variance of a set of fit results.

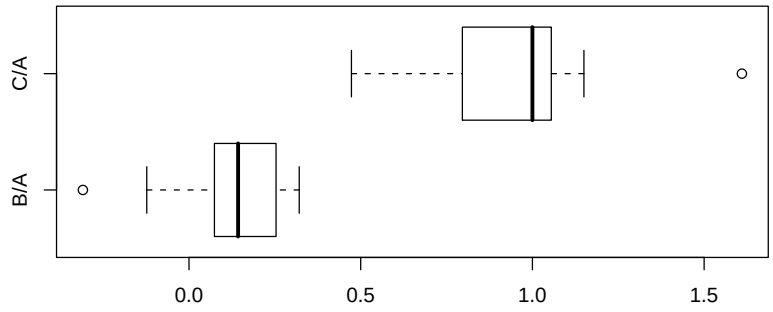


FIGURE 10.6: Box-plot of the all-real water ratios.

Water Source	Value	$\bar{x}$	σ
2	$\frac{B}{A}$	0.279	0.0286
	$\frac{C}{A}$	0.710	0.220
3	$\frac{B}{A}$	0.162	0.0312
	$\frac{C}{A}$	0.867	0.103
Combined Data	$\frac{B}{A}$	0.229	0.0652
	$\frac{C}{A}$	0.778	0.196
Frey 2001	$\frac{B}{A}$	0.266	0.01
	$\frac{C}{A}$	0.722	0.007

TABLE 10.6: Water SHG polarisation fit ratio summary.
For fits with real parameters.

10.4 Revisiting the R6G data

In light of the findings about the repeatability of the water samples on the SHG experiment, the repeatability of the R6G data should be reconsidered.

It was found with both pure water and aqueous R6G solutions that whilst qualitatively similar data could be collected over multiple experiments, there was some variation in the fit co-efficients. The importance of cleanliness was highlighted, as was the possible variation in the purity of water provided by different "pure" water sources; when contamination was apparent (for example by the sample holder not being thoroughly cleaned) the polarisation curves are qualitatively quite different to the modal water polarisation curves, often requiring complex parameters when fitting the data to the theoretical model. "Water" samples showing such deviations were isolated and discarded.

These differences are harder to spot when dealing with a solution; one is less sure if the behaviour is authentic or the result of contamination.

The overall conclusion was that the experiment can provide results for simple systems that are as repeatable as is seen for lower concentration R6G samples (i.e. that there is a similar level of variability in the coefficients). It is therefore reasonable to look beyond experimental inconsistencies when explaining the data for solutions with concentrations around 1×10^{-5} mol dm^{-3}. The concept of an unstable surface is appealing, where it may be that either the surface is in a state of flux, or that there are "clumps" of behaviour and that these drift around, varying the environment under the laser's focus. In either case, the long timescale of SHG measurements in this experiment (in the order of 50 seconds per data point, or 45 minutes for a complete polarisation scan) dwarfs the likely speed of change at the surface; in the "clump" system proposed these clumps would likely be moving on the same timescale as particles on the surface, analogous to the every day experience of watching small fragments of fruit moving in a cocktail which suggests that this may be comparably fast. This also helps to explain why stability experiments (making SHG measurements with all the controlled parameters fixed) shows the system to be relatively stable – the instabilities in question are being swamped out by the 50 second (or 1000 point) averaging needed to get good signal to noise behaviour. To improve this, one would need a system which allows this experiment timescale to be greatly reduced, for example by improving the signal to noise in the detector system, or using a laser source with a higher pulse rate, thus either needing either less points, or collecting the same number of data points more rapidly.

Chapter 11

The value of eScience to Interfacial Chemistry

The work presented in this thesis was inspired, in part, by the efforts of Frey, schraefel *et al.* in their "smartLab" work analysing the behaviour of the organic chemist in the laboratory[18]. Their work highlighted the semi-structured nature of synthetic organic chemistry. They found that a chemist will start work with a clear plan of the experiment being undertaken, but that as the experiment progresses the chemist must be prepared to deviate from this as they discover new aspects of their work. The eScience solution to support this work must, therefore, be flexible enough to accommodate this, whilst maintaining a record of both the originally planned work, and that which actually occurred. The pervasive nature of the solution provided by Frey *et al.* and the ability to record "free-form" sketches, were major contributors to their success. The chemists involved in the study were also impressed with the ability to recall and access their laboratory work in a number of ways, both in the lab and away from it. In this work, the experimental work-flow is more rigid, allowing some of the flexibility of the "smartLab" solution to be removed; however, the ability to record notes as an experiment progresses was still considered to be very important. In addition, it was thought that the SHG experiment was likely to be more sensitive to changes in the laboratory environment than had been encountered in the work of Frey *et al.* The instantaneous feedback of critical events provided by the implementation, saved time on multiple occasions as the experimenter was informed of these events as they occurred and hence could take mitigating actions. Previously, these events would have gone unnoticed until some time later, wasting the elapsed experimental time. The ability to review the exact conditions under which an experiment was performed also promotes the credibility of the work, giving a third-party observer a greater understanding of the experiment. A significant link between poor experimental data and environment events has yet to be found, although it maybe that this link is being masked by the general inconclusiveness of the SHG data collected for the compounds examined here, it being difficult to isolate fault conditions when there are

so few robust data-sets to benchmark against. Subsequent work with other compounds has demonstrated more robust data, and it is expected that a future publication will report a more conclusive link.

Of far greater impact to this work was the recall of experimental data. With the eScience solution in place, experimental data were easily searched and retrieved through a web interface, along with a preliminary, automated, analysis. This meant that fits were attempted for all data collected; previously only those data which appeared to show promise would have been fitted. This automated data management has also laid the foundations for automatic publication and sharing of data. The data collected during this work were made available within the author's research group by means of a web-based "blog" interface, which allowed them to be viewed and discussed with others, without the need to schedule face to face meetings, a process which has become increasingly difficult with the workload of the modern academic. With some regret, this technology was not so fully exploited in this work as it could have been; subsequent work has been using the same infrastructure to much greater effect. The ability to publish these data electronically and automatically will allow for much greater access to, and validation of, scientific output. As discussed in chapter 3, publication of all the data that support a result allows other parties both to confirm a result and also to reuse the data in other work. These concepts are only briefly considered for the case of SHG; however, the author is aware of other projects in which data are automatically released to the wider community after an embargo period. Having also been inspired by the early smartLab work, the UK X-Ray Crystallography service now releases all data collected after such an embargo, unless there are good reasons related to intellectual property not to do so. The embargo period is necessary as it allows the originating researcher time to publish their findings and gain the deserved recognition for their work. However, the automated sharing thereafter prevents data being lost if the originator decides against publication. In contrast to traditional X-Ray crystallography publishing, all the raw data that are collected when determining the structure are published, providing the reader with the ability to both gain a greater understanding of the sample and to reanalyse the data should they wish to.

The work described here has provided the foundation for a handful of other projects. The use of broker technologies to support laboratory data management has proved itself to be reliable and extensible, and has since been implemented in a number of other scenarios, ranging from a biochemistry laboratory to high-energy physics experiments. In more closely related work, the SHG database designed here is still being used to handle data for the Southampton SHG experiment, providing the data repository for a subsequent development of the Blog tools discussed previously. The data collection and review tools helped to highlight a number of ways that the Southampton SHG experiment could be improved, and these issues are being addressed in subsequent work.

Chapter 12

Future work

Whilst performing the research and subsequent analysis for this work, a number of points have been identified that could improve the efficiency, accuracy or reliability of the experiment. This work only forms a small part of the entire picture for an ideal "publish@source" system, hence some consideration must be given to expanding this work into other areas. The author is aware that during the preparation of this document other researchers in the School of Chemistry have continued this research area, and that some of these items have already been addressed as part of other work.

SHG rig improvements

There is some scope for improvements to the apparatus beyond that which is considered maintenance. At the end of this work, only some of the variable optics were under computer control, requiring periodic human intervention to complete data collection. Automation of these optical items would improve the efficiency of the apparatus, allowing the experiment to complete without human intervention. This will also allow some improvement in accuracy to be gained. Firstly, if indexible rotators were used on all automated optics, then the repeatability of the experiment would improve. Secondly, by automating the power and monochromator settings, data collection dependent on these factors would be made easier, and hence scans could be expanded. This would again increase the reliability of data collected as more experiments would be performed under optimum conditions.

When considering the review of data, it has been suggested that there would be much merit in collecting other, more general, data around the apparatus during an experiment. Still cameras could be mounted above the optics bench and in front of the electronics rack, which would periodically take images during experiments or times when the lab is in use; a secondary record of the configuration and settings would therefore created,

independent of the SHG experiment database. This may help fill in details not automatically recorded, such as which sample holder was in use, was a lid placed over the sample or were the black-out covers in place over the optics. It would also provide the ability to check if the electronics settings recorded in the database are correct.

During this work, the installation of RFID technology into the lab space was started as part of another project which never completed. This was going to provide logging of the location of certain items, such as the sample holders, around the lab. In the case of sample holders, RFID readers were to be placed in key positions, and a "read" of the RFID tag to be performed at key times; for example when the holder is placed on the laser table, or when it was cleaned (possibly with multiple readers to allow indication of which cleaning method was performed). Implementation of such a system would provide an additional strand of information for the experiment provenance chain, allowing automated recording of this information.

For more in-depth study of aqueous systems it would be advantageous for there to be a dedicated water source, under the direct control of those researchers performing the SHG experiment.

Publish@Source completion

SHG Experiment

To complete the specific publish at source system for the Southampton SHG experiment, a few areas need addressing. Firstly, a system to record the analysis and post processing of the data needs implementing; at a minimum, recording the choices that are made when selecting which data to publish and any additional processing that was needed beyond that which forms part of the published analysis procedure. Secondly, some consideration needs to be given to the assignment of access rights and permissions throughout the system. Work has been undertaken by the computer science community in this field, much of which is applicable to this work; a solution based on existing technologies should be reasonably trivial to implement. The existing use of web-server based processing for much of the presentation and publishing of data should make this easier. Some thought also needs to be given to the long-term curation of the raw and derived databases created as part of this work. This would be especially important if any publications rely on live access to them to produce their supporting data. In order to integrate with existing electronic curation (for example the University of Southampton ePrints service) an appropriate method may be to produce "mini publications" that contain the raw data for a specific experiment, which can then be referenced as source material from a formal publication. A version of this thesis, to be included on the supplementary data disk and in the University ePrints, will be an example of this; pages from the recall tool described

in section 8.2.5 are stored as HTML files, which are linked to the appropriate text and images in a PDF version of the main document. However the work to create these links will be performed manually and automation of this process would be attractive.

Other Experiments

When looking to other experiments based on the experience of the SHG experiment, a difficult question is posed. How specific should the implementation be? Should a specific data storage solution be developed for each experiment, or should a general database be developed, and the experiment specificity be held within the data stored? The second of these is certainly attractive if one is wishing to provide a solution to the widest market, as it allows for experiments to be added to a system with minimal work on the back-end. However, one has to consider how to present the data on retrieval, one possible solution is to include a preview image/text with the data which will display with only modest software requirements (for example requiring only a web-browser) with the full experiment specific data files also available for those people with the appropriate tools to view them. It can be seen that such a system has more in common with a meta-data annotated file store, such as the DICE Storage Resource Broker[84], than with the database developed for SHG. The reuse of other technologies is attractive as it can reduce the amount of development work required.

One needs to consider how the data are captured into such a system. Firstly, the data need to be available to a networked computer, many new instruments are controlled by commodity personal computers, so this should be quite trivial to achieve. In the case of older apparatus, the device may have some form of external storage to allow export of data in a computer readable format, in which case the data can be transferred to a computer workstation and then uploaded; giving consideration to the availability of the technology required (for example being able to read old format computer media). If this is not an option, then capture of the hard-copy output into the digital domain needs to occur and here optical scanners and digital cameras can be useful. Once the data are available electronically, data handling can be considered in one of two ways. Where an apparatus exists with a commercial control system, or with files collected from an older apparatus, the data injection step is a file upload, this may be a manual operation by which the user provides meta-data at the same time as inserting the file, or maybe an automated system which grabs all files from a directory, uploading them to the system for later tagging and assignment to an experiment. Alternatively, where a custom built control system exists (such as for the SHG experiment), then data upload can be built into the apparatus control and meta-data collected as part of the experiment setup, maybe even automatically if the sample can be tagged in some computer readable way.

Once the data have been captured into an electronic storage system, tools must be provided to allow for comment, post processing and analysis. During this work, the power

of a blog as a comment tool was considered and found to be useful (the implementation of the blog was the focus of another research project, and this work was used as one of a number of test cases). The ability to comment on and share the data in a transparent fashion was most impressive. Including data from the experiment into existing collaborative systems can provide the interested parties with the ability to discuss the data using familiar tools. Other tools that could be considered include traditional desktop environments (such as Lotus Notes or Microsoft SharePoint), web based environments (blogs have been considered, but also worth investigation are wikis and social networking tools), real-time discussion tools (for example, instant messaging and shared desktop tools) and immersive environments (for example Second Life, Quaq). Also for consideration should be the parallel access by multiple collaboration tools, for example that comments posted in the blog should appear in the Second Life interface and vice-versa.

Part of the capturing of comments should also consider capture of the data analysis. This is especially important when multiple steps occur; a record of, and the resulting data at, each step must be available for subsequent inspection. Technologies to support this need to be developed, taking into account that different experiments and indeed different users will have their own preferences for data analysis.

Finally, tools need to be provided to view the whole package of stored data for an experiment, in one aspect this will be provided through the discussion tools mentioned, but tools for review post publication need to be considered. In this case, it is likely to be read-only and with the ability to filter out those internal discussions that may not be suitable for publication at a specific time. For the foreseeable future, traditional publication will be an important part of academia, and the production of high quality graphics and plots for inclusion in such papers should be considered, either automatically within the experiment data system, or externally with an easy data export facility.

Appendix A

Glossary

The purpose of this section is to provide a quick reference to aid the understanding of some areas of this thesis, and, where appropriate, to provide external reference for more in depth understanding.

Some of the glossary entries are verbatim copies, modified copies, or extracts from http://www.wikipedia.org and, as such, are referenced on a per entry basis. Entries referenced in this way are covered under the GNU Free Documentation License, the full text of which is available at http://www.gnu.org/licenses/fdl.html, under the terms of which they should be considered as a combined work. All other entries are my own work, or are referenced appropriately, in line with University of Southampton guidelines.

A.1 Chemistry Terms

1. **Aqueous Solution**
   A solution involving water as the predominant solvent.

2. **Centrosymmetric**
   Having a symmetrical arrangement of radiating parts about a central point.
   Taken from http://www.hyperdictionary.com/dictionary/centrosymmetric

3. **Coherent / Coherence**
   Coherence is a property of waves that measures the ability of the waves to interfere with each other. Two waves that are coherent can be combined to produce an unmoving distribution of constructive and destructive interference (a visible interference pattern). Waves that are incoherent, when combined, produce rapidly moving areas of constructive and destructive interference and therefore do not produce a visible interference pattern.
   Taken from http://en.wikipedia.org/wiki/Coherence_(physics)

4. **Collimated Beam**
   Collimated light beams are those in which the light rays are perfectly parallel. Light from a laser source is often almost collimated. Light from the sun and distant stars is collimated to a good approximation due to their distance from us.

5. **Dipole / Dipole Moment**
   A dipole is a pair of electric charges or magnetic poles of equal magnitude but opposite polarity, separated by some (usually small) distance. Dipoles can be characterised by their dipole moment, a vector quantity with a magnitude equal to the product of the charge or magnetic strength of one of the poles and the distance separating the two poles. The direction of the dipole moment corresponds to the direction from the negative to the positive charge or from the south to the north pole.
   Taken from, and for more information see http://en.wikipedia.org/wiki/Dipole_moment

6. **Dielectric Medium / Dielectric Material.**
   Most generally, a dielectric is an insulator, a substance that is highly resistant to flow of electric current. The molecules in a dielectric strongly resist allowing a flow of electrons, and ionising, preferentially forming dipoles.

7. **Electric Dipole Approximation**
   A light wave consists of electric and magnetic fields which vary sinusoidally at high frequencies. When light propagates through matter it will therefore induce motion of the charged particles that constitute the material. In a dielectric medium the charges are bound together and will start to oscillate in the applied electric field; they form oscillating electric dipoles. The contribution from the magnetic field part of the light and from electric quadrupoles is much weaker and is usually neglected. This is the electric-dipole approximation.

8. **Electric Susceptibility**
   The electric susceptibility χ_e is essentially the constant of proportionality (which may be a tensor) relating an applied electric field, $\mathbf{E}$, to the induced dielectric polarisation, $\mathbf{P}$, such that:
   $$\mathbf{P} = \epsilon_0 \chi_e \mathbf{E}$$
   Where ϵ_0 is the electric permittivity of free space.
   Taken from http://en.wikipedia.org/wiki/Electric_susceptibility

9. **Fresnel Equations**
   The Fresnel Equations describe the behaviour of a light beam when moving between two media of differing refractive index. At the point the beam meets the interface, reflection and/or refraction may occur. Figure 9 shows this schematically.
   The fraction of the incident light that is reflected from the interface is given by the

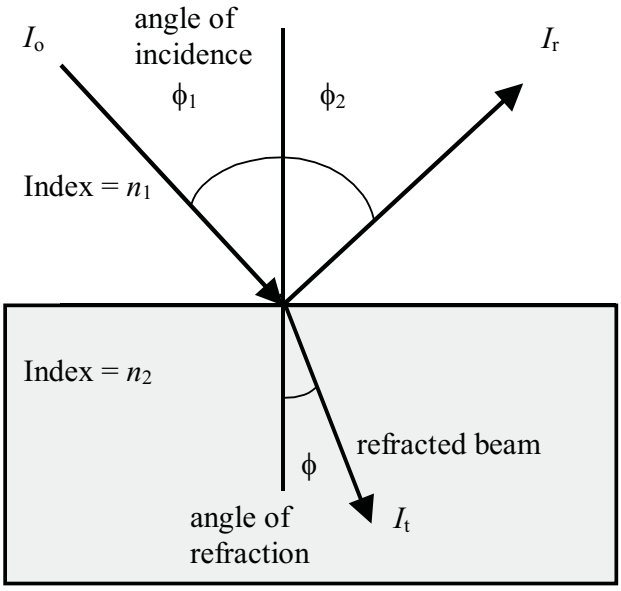


FIGURE A.1: Schematic of light reflecting and refracting from an interface between materials of refractive index n_1 and n_2.[66]

reflection coefficient R, and the fraction refracted by the transmission coefficient T. The Fresnel equations may be used to calculate R and T in a given situation. The intensity of the reflected (or transmitted) light is dependent on the polarisation of the incident light. For light with an electric field perpendicular to the plane of figure 9 (s-polarised), the reflection co-efficient is

$$R_s = \left(\frac{n_1 \cos(\phi_1) - n_2 \cos(\phi)}{n_1 \cos(\phi_1) + n_2 \cos(\phi)} \right)^2$$

and for light polarised in the plane of figure 9, the reflection co-efficient is

$$R_p = \left(\frac{n_2 \cos(\phi_1) - n_1 \cos(\phi)}{n_2 \cos(\phi_1) + n_1 \cos(\phi)} \right)^2$$

The transmission co-efficients are then found by

$$T_p = 1 - R_p$$

and

$$T_s = 1 - R_s$$

10. **Isotropic / Anisotropic**
    Isotropic means "independent of direction".
    Anisotropic is usually used to describe a directionally dependent phenomenon.

11. **Interface / Bulk Material**
    Bulk material refers to regions of a material that are away from its edges, and hence it can be considered that the only interactions are those coming from other molecules of the same materials, and in the case of liquids and gases these forces can be considered to be symmetrical.
    Interface regions are those where two materials join, hence interactions from two materials need to be considered. Interfaces tend to be sharp transitions from one material to the other (over the distance of a few molecules), hence the molecules at this transition will have differing forces acting on their opposing sides.

12. **Laser**
    A Laser (light amplification by stimulated emission of radiation) is a device which uses a quantum mechanical effect, stimulated emission, to generate a coherent beam of light. Light from a laser is often very collimated and monochromatic, but this is not true of all laser types.
    From http://en.wikipedia.org/wiki/Laser

13. **Monochromatic/Monochromaticity**
    Light of only one wavelength, or referring to a source that produces light of only one wavelength, such as a laser.

14. **Non-Linear Effects**
    Generally, non-linear effects are those which cannot be modelled by a linear relationship.
    Nonlinear optics is the branch of optics which describes the behaviour of light in nonlinear media, that is media in which the polarisation, **P**, responds non-linearly to the electric field, **E**, of the light. This non-linearity is typically only observed at very high light intensities such as provided by pulsed lasers.
    For further information see http://en.wikipedia.org/wiki/Non-linear_optics

15. **Nature of Light**
    Light can be considered in two ways, it can be thought of as an electromagnetic waveform, or quantum mechanically as photons. For the purpose of this work, the waveform notation is more convenient.
    When considering light as a waveform, it is thought of as a transverse electromagnetic wave consisting of an electric field and a magnetic field whose field vectors are perpendicular to each other and to the direction of propagation. Electromagnetic radiation is represented as a sinusoidal harmonic wave, which is characterised by its frequency, phase and propagation direction. To fully characterise an electromagnetic wave, the direction of its electric and magnetic components relative to the direction of propagation needs to be known, it is this relative orientation that defines the light's polarisation.

16. **Permittivity of free space**

In electromagnetism, permittivity ϵ is a measure of how much a medium changes to absorb energy when subject to an electric field. The permittivity ϵ and magnetic permeability μ of a medium together determine the velocity v of electromagnetic radiation through that medium:

$$\epsilon\mu = \frac{1}{v^2}$$

which in a vacuum becomes:

$$\epsilon_0\mu_0 = \frac{1}{c^2}$$

linking the permittivity of free space, ϵ_0, magnetic permeability of free space, μ_0, to the speed of light in a vacuum, c.

Taken from http://en.wikipedia.org/wiki/Permittivity_of_free_space

17. **Polarisation**

As defined in the nature of light section (entry 15), the polarisation of light is defined by the direction of its magnetic and electric vectors relative to the vector of propagation. For simplicity, considering only the electric field vector (the position of the magnetic field vector follows from it), the polarisation is defined as follows. For 'ordinary' unpolarised light, the electric field vector will rotate and change amplitude arbitrarily with respect to the propagation vector. If the relative orientation of the electric and propagation vectors is constant, then the light is referred to as being plane or linearly polarised. When a plane polarised light wave is polarised in the plane of incidence, it is defined as being P polarised, and when it is perpendicular to the plane of incidence, it is defined as being S polarised.

Circularly and elliptically polarised light can be thought of as the superposition of two harmonic, co-linear, linearly polarised light waves of the same frequency. When the two waves are in phase and of the same amplitude, this is the degenerate case and is considered as plane polarised light. When there is a phase difference of 90^o, and the amplitudes are equivalent, the light is said to be circularly polarised. Also, other combinations are defined as being elliptically polarised.

18. **Refractive Index**

The refractive index, n, of a material at a particular frequency is the factor by which electromagnetic radiation of that frequency is slowed down (relative to vacuum) when it travels inside the material.

For a non-magnetic material, the square of the refractive index is the dielectric constant. For a general material, it is given by

$$n = \sqrt{\epsilon\mu_0}$$

where μ_0 is the permeability of free space, and ϵ the dielectric constant.
See also entry 16, Permittivity of free space.
Taken from http://en.wikipedia.org/wiki/Refractive_index

A.2 Computer Science Terms

1. **Agent**
   In computer science, a software agent is a piece of autonomous, or semi-autonomous computer software. It is programmed to perform a (range of) simple tasks, often to aid the movement of data from one system to another. Many individual communicative software agents may form a multi-agent system.
   From http://en.wikipedia.org/wiki/Software_agent

2. **Analogy Based Design Process**
   When working with personnel from another speciality, communication of design ideas and requirements can become a problem. By using a common analogy (possibly from outside both fields of expertise) these problems can be overcome.
   For example, Schraefel *et al*[18] used the analogy of making tea to model a chemists use of a laboratory notebook when designing an electronic replacement.

3. **Broker**
   In commerce, a broker is a party that mediates between a buyer of shares and a seller of shares. A brokerage is a firm that acts as a broker.
   A computer science broker is analogous to this. Its task is to connect data providers to data subscribers, often merging data streams from many providers, and filtering the results to suit specific data subscribers.

4. **Client/Server Model**
   A client-server architecture is a network architecture in which each computer or process on the network is either a client or a server. Servers are powerful computers or processes dedicated to managing disk drives (file servers), printers (print servers), or network traffic (network servers). Clients are PCs or workstations on which users run applications. Clients rely on servers for resources, such as files, devices, and even processing power. This architecture is primarily distinguished from other architectural styles by its reliance on structuring the system using layers.
   From http://en.wikipedia.org/wiki/Client_server

5. **Data Acquisition Card / Data Acquisition device**
   A device, or computer expansion card that allows the monitoring of physical data. This will typically involve an analogue to digital conversion. Toggle switches, however, can be connected directly to digital inputs, as has been done for some of the sensors in the smartLab.

6. **Database**
   A database is an information set with a regular structure. Its front-end allows data access, searching and sorting routines. Its back-end affords data inputting and updating. A database is usually but not necessarily stored in some machine-readable format accessed by a computer. There are a wide variety of databases, from simple tables stored in a single file to very large databases with many millions of records.
   From http://en.wikipedia.org/wiki/Database

7. **Firewall / Stateful Firewall / Packet Filtering Firewall**
   In the construction industry, a firewall is a windowless, fireproof wall built to prevent fire from spreading beyond one section of a building. The computing world use is analogous to this, there the term firewall is used for a piece of hardware or software put on the network to prevent some communications forbidden by the network policy.
   From http://en.wikipedia.org/wiki/Firewall

8. **LAN**
   A local area network (LAN) is a computer network covering a local area, such as a home, office or small group of buildings like a college. It is typically used to allow collaboration, and sharing of resources between a number of machines. This could include file storage, printers and access to outside networks such as the Internet.

9. **Linux**
   Linux is the name of a computer operating system and its kernel. It is the most famous example of free software and of open-source development.
   Strictly, the name Linux refers only to the Linux kernel, but it is commonly used to describe entire Unix-like operating systems (also known as GNU/Linux) that are based on the Linux kernel and libraries and tools from the GNU project.
   From http://en.wikipedia.org/wiki/Linux

10. **Meta-Data**
    Meta-data is data about data. A good example of this is a library index system, which describes where the data is and it's nature, but doesn't replicate it's content. Meta-data is often used on the World Wide Web to provide information about pages in a computer readable format, such as keywords, author and publishing tool.
    See http://en.wikipedia.org/wiki/Metadata for further information

11. **Multicast / Unicast / Broadcast**
    Multicast is the transmission of the same data to a specific number of clients simultaneously using a mechanism that allows the originator to send the data only once. By comparison with multicast, conventional point-to-single-point delivery is called unicast, whereas delivery to every node of the network is broadcast.

12. **Online / Offline (network applications)**
    Networked application are those which interact with some networked resource to perform their core duties. An online application is one that is in constant contact with its networked resource, where as an offline application is one which stores enough data locally to perform its tasks and then periodically synchronises with the networked resource.

13. **"Pervasive Computing" / Ubiquitous computing**
    Ubiquitous computing is a term describing the concept of integrating computation into the environment, rather than having computers which are distinct objects. Promoters of this idea hope that embedding computation into the environment would enable people to move around and interact with computers more naturally than they currently do. One of the goals of ubiquitous computing is to enable devices to sense changes in their environment and to automatically adapt and act based on these changes based on user needs and preferences.
    From http://en.wikipedia.org/wiki/Ubiquitous_computing

14. **Schema (Database)**
    The word schema comes from the Greek word meaning shape or more generally plan. In computer science, a schema is generally a model.
    With respect to databases, a schema is a description of the databases structure, or a sub-part thereof.

15. **Script / Scripting Language**
    A programming language typically used for simple tasks such as data translation or simple manipulations. Scripting languages are typically "run-time" compiled, which means that they are converted to machine code as they run, not in advance as with many programming languages. Common uses for scripting languages include running common sequences of programs or producing dynamic web pages.

16. **Tablet PC**
    A small portable computer, typically presented as a screen with a stylus to allow data entry. Tablet PCs don't usually have keyboards or mice, but these can be added allowing the tablet to perform like a laptop computer. They are typically used to replace paper notebooks, as they can provide a similar interface.

17. **TCP/IP Network**
    A computer network using TCP/IP as it's logical transport layer. TCP/IP is a hardware independent transport layer, that can be mapped onto a wide range of physical transports. It's most common uses today are as the underlying protocol of the global Internet, and as a transport layer in may LAN's.
    See http://en.wikipedia.org/wiki/TCP/IP for more details

18. **Terminal**

    (a) A client computer, typically with little local processing power, user to access a remote system.

    (b) Software running on a Personal computer used to connect to a remote system.

19. **TTL**
    Transistor-Transistor Logic. Signalling system where binary 0 is between 0 V and 0.8 V, and binary 1 is between 2 V and 5 V. It is notable as being the first widespread digital signalling standard.
    For further details see http://en.wikipedia.org/wiki/Transistor-transistor_logic

20. **UNIX Time-stamp**
    Time in seconds since Midnight, January 1st 1970.

21. **URL**
    A Uniform Resource Locator, URL (sometimes pronounced as "earl" or more often spelt out), or web address, is a standardised address for some resource (such as a document or image) on the Internet. First created by Tim Berners-Lee for use on the World Wide Web, the currently used forms are detailed by IETF standard RFC 2396 (1998).
    For further details see http://en.wikipedia.org/wiki/URL

22. **Window (WRT Graphical Interfaces)**
    A window is a graphic, usually rectangular in shape, containing either some kind of graphical interface, or a textual representation, of the output of and allowing input for one of a number of simultaneously running computer processes.
    In graphical user interfaces windows are suggested to be objects (like papers or books) on a desktop: when two overlap, one is on top of the other, with the overlapping part of the lower window not shown; when moving, minimising, resizing, or closing the upper window, hidden parts of the lower window reappear.
    Taken from http://en.wikipedia.org/wiki/Window_(computing)

23. **XML**
    Extensible Markup Language (XML) is a markup language used for creating special purpose markup languages. A markup language is one where data is separated by markers to indicate their purpose of formatting. XML uses angle brackets < and > to delimit markers from data, and markers exist in pairs, to delimit the ends of the data, e.g. <Marker>Data</Marker>.
    For more details see http://en.wikipedia.org/wiki/XML

Appendix B

Supporting Data CD

The supporting data cd, attached, contains electronic versions of this thesis, publication quality copies of the images and plots used, an archive of the raw data generated during this work, and an archive of the software produced.

The CD also includes a HTML interface. On a windows PC this should automaticaly load when the CD is inserted. If this doesn't occur, or on other operating systems, please view the README.HTML file in the CD root directory.

If the attached CD is damaged or corrupted, an electronic version can be found stored in the University of Southampton ePrints system, which can be accessed using the web URL http://eprints.soton.ac.uk/64884

B.1 CD Structure

```
Directory Contents.

/thesis/
   PDF and TeX copies of the thesis, and EPS copies of images

/data/
   Archives of the raw data files, database dumps and a static
   copy of the experiment review website.

/HTML/
   Pages and build tree for the CD interface

/software/location
   Location awareness programs
```

```
/software/location/bluelocate
   Bluetooth beacon database logger

/software/location/bluelocateDB
   Location awareness archive database

/software/location/bluemonitor
   Bluetooth beacon monitoring agent

/software/monitoring
   Laboratory monitoring software

/software/monitoring/BackupAgent
   Agent which stores topics from the broker environment streams into
   a PostgreSQL Database.
   The folder includes database schema and database dump from this
   work as an example.

/software/monitoring/mimicAgent
   Repeats temperature data on 1 second timebase for MQ charting client

/software/monitoring/MQAgents_Binary
   Binaries of the above agents for linux

/software/monitoring/original-WWW-Recall
   Early web based environment recall tool. Very slow on big databases

/software/monitoring/PublishLab
   Early lab environment client, based on National Instruments LAB-PC+
   ADC card.

/software/monitoring/PublishLab_Binary
   Windows binary of PublishLab

/software/monitoring/UIBer
   IBM dashboard client message translator

/software/remote_control
   Lego microscope software
```

```
/software/remote_control/lego_mqtt/intelscope
   Microscope camera control MQ bridge

/software/remote_control/lego_mqtt/jr-RCX
   Lego mindstorms-mqtt bridge

/software/remote_control/lego_www
   Lego microscope control webpages

/software/SHG_Expt/database
   Schema and database dump for the SHG database

/software/SHG_review/
   CombeChem SHG Experiment database tools

  /software/SHG_review/autoBlog
   The MQTT listening end of experiment blog entry generator

  /software/SHG_review/database
   Experiment review database schema and dump

  /software/SHG_review/expt_extract
   Experiment database -> review database, and automatic fit tools

  /software/SHG_review/www/expt_review
   Web frontend to the review database

/software/SHG_review/www/smartSHG
   Web front-end to the experiment database
```

B.2 Software Manuals

The software directory contains much of the software and scripts produced as part of this work. Each package contains a "readme" file, these are included here for convenience.

B.2.1 CombeChem MQTT Database Backup Agent

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

This code is based on, and includes redistributables from, IBM IA93 -
"WebSphere MQ Telemetry Transport C language implementation" and may
contain additional licensing restrictions.  Please refer to included
IA93_Lic_en.txt

============================
Dependencies

UnixODBC (most recently built against version 2.2.11)
IBM IA93 -- MQTT Sample Library C (no longer available)
Suitable C build environment (eg gcc, GNU make, GNU ld)
BrokerBackup database

============================
Database Setup

Create database using the structure.sql file in the database directory.
Create a user with Write and Select privileges to this database.
Create a system ODBC DSN for this user.

============================
Build Instructions

Copy the following files from the IA93 distribution to the build
directory.
MQIsdp.h
mspfp.h
mspfp.c
libwmqtt.so
```

```
Enter the source directory, and use the Makefile

to build
$ make clean
$ make

To test (using the configuration in the Makefile)
$make run

============================
Configuration.

Broker and topic are configurable at runtime (as below).
Database Server (type, host, port, user, host, database, table) configured
as a system ODBC source.
Other configuration to be done at source build time.

./<Excutable> wmqtt://<clientname>@<broker host>:<broker port>/
   <topic(s) to subscribe>

eg
./MQTempSens \\
wmqtt://MQReadTemp@smartlab.combe.chem.soton.ac.uk:1883//SmartLab/SHG/Env/#

============================
Source Configurable Parameters.

   Database Server (odbcsource, username, password).
 Configuration is in MQRead.c, just below the library includes.
```

B.2.2 CombeChem MQTT Mimic Agent

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

This code is based on, and includes, redistributables from, IBM IA93 -
"WebSphere MQ Telemetry Transport C language implementation" and be subject
to additional licensing restrictions.  Please refer to included IA93_Lic_en.txt

================================================
Dependencies

IBM IA93 -- MQTT Sample Library C
Suitable C build environment (eg gcc, Gnu Make, Gnu Ld)

================================================
Build Instructions

Copy the following files from the IA93 distribution to the build directory.
MQIsdp.h
mspfp.h
mspfp.c
libwmqtt.so

Enter the source directory, and use the Makefile

to build
$ make clean
$ make

to test (using the configuration in the Makefile)
$ make run
```

```
================================================
Configuration.

Only the broker and topic are configurable at runtime (as below).  Other
configuration to be done at source build time.

./<Excutable> wmqtt://<clientname>@<broker host>:<broker port>/
 <topic(s) to subscribe>

eg
./MQTempSens \\
wmqtt://MQReadTemp@smartlab.combe.chem.soton.ac.uk:1883//SmartLab/SHG/Env/#

================================================
Source Configurable Parameters.

(sub)Topics to mimic, and mimic destination configured in the array just
above the main function definition in MQRead.c
```

B.2.3 CombeChem MQTT Agents - Linux Binary

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

This code is based on, and includes redistributables from, IBM IA93 -
"WebSphere MQ Telemetry Transport C language implementation" and may
be subject to additional licensing restrictions.  Please refer to included
IA92_Lic_en.txt

============================
Dependencies

UnixODBC (last test build against version 2.2.11) (and database configured
   to match the source code)
libwmqtt.so from IBM IA93 -- MQTT Sample Library C

============================
Execution Instructions.

Copy libwmqtt.so from the IA93 distribution to the directory containing
the binaries.

Execute the binaries (MQBackup and MQTempSens) from the command
line, with broker and topic specified as given in the following examples,
MQ wildcards are allowed.

./<Excutable> wmqtt://<clientname>@<broker host>:<broker port>/
<topic(s) to subscribe>

./MQBackup  \\
wmqtt://MQReadComp@smartlab.combe.chem.soton.ac.uk:1883//SmartLab/\+/Env/#
./MQTempSens \\
wmqtt://MQReadTemp@smartlab.combe.chem.soton.ac.uk:1883//SmartLab/SHG/Env/#
```

```
============================
Configuration.

Broker and topic are configurable at runtime (as above).
Database Server (type, host, port, user, host, database, table)
configured as a system ODBC source.
Other configuration to be done at source build time.

============================
Source Configurable Parameters.

MQBackup
   Database Server (odbcsource, username and password)

MQTempSens
   (sub)Topics to mimic, and mimic destination.
```

B.2.4 CombeChem laboratory monitoring recall tool - Web

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

============================
Dependencies

Webserver (tested on apache2)
PHP installation (tested on PHP5)
Database for BrokerBackup agent (tested with postgresql)
PHPChart (GPL code, included)

============================
Install Instructions

Copy the files to a directory on the webserver.

Edit labrecall.php to include the correct ODBC source, username and
password in the odbc_connect line.

Access labrecall.php using your web-browser through the webserver.
```

B.2.5 CombeChem MQTT NI-DAQ Windows capture agent

```
Program to capture data using a National Instruments LAB-PC+ ADC card,
and publish message via MQTT
================================
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

This code is based on, and includes redistributables from, IBM IA93 -
"WebSphere MQ Telemetry Transport C language implementation" and may contain
additional licensing restrictions.  Please refer to included IA92_Lic_en.txt

This code includes portions of, and redistributables from, National
Instruments Traditional NI-DAQ Implementation. Please refer to included
NIDAQ-lic-en.rtf.

============================
Dependencies

National Instruments Traditional NI-DAQ 6.9.2 (available on CD in
   /software/3rdPartyLibraries)
IBM IA93 -- MQTT Sample Library C
Suitable C build environment (last built with MS Visual C 6.0)

============================
Build Instructions

Copy the following files from the IA93 distribution to the Include
directory.
   MQIsdp.h
   mspfp.h
   mspfp.c

Copy wmqtt.lib from the IA93 distribution to the lib directory
Copy wmqtt.dll from the IA93 distribution to the Debug directory
```

```
Load the project file "PublicLab.dsw" into MS Visual C IDE, use the
included build tools.

============================
Configuration.

Broker and topic are configurable at runtime (as below).
Other configuration to be done at source build time.

./<Excutable> wmptt://<clientname>@<broker host>:<broker port>/
   <topic(s) to subscribe>


============================
Source Configurable Parameters.

  Capture card details, and sensor parameters
 Various variables defined at the top of the main function in src/publish.c
```

B.2.6 CombeChem MQTT NI-DAQ Windows capture agent - BINARY

```
Program to capture data using a National Instruments LAB-PC+ ADC card,
and publish message via MQTT
============================
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

This code is based on, and includes redistributables from, IBM IA93 -
"WebSphere MQ Telemetry Transport C language implementation" and may contain
additional licensing restrictions.  Please refer to included IA92_Lic_en.txt

This code includes portions of, and redistributables from, National
Instruments Traditional NI-DAQ Implementation. Please refer to included
NIDAQ-lic-en.rtf.

============================
Dependencies

Installed NI-DAQ drivers (including libraries)
wmqtt.dll from IBM IA93 -- MQTT Sample Library C
============================

Execution Instructions.

Copy wmqtt.dll from the IA93 distribution to the directory containing
the binaries.

Execute the binary from command prompt, with broker and topic specified
as given in the following examples, MQ wildcards allowed.

./<Excutable> wmptt://<clientname>@<broker host>:<broker port>/
<topic(s) to subscribe>
```

```
============================
Configuration.

Broker and topic are configurable at runtime (as above).

Other configuration to be done at source build time.

============================
Source Configurable Parameters.

    Capture card details, and sensor parameters
    Various variables defined at the top of the main function in src/publish.c
```

B.2.7 CombeChem bluetooth device tracker

```
(c) University of Southampton 2007
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

================================================
Dependencies

Perl (tested with version 5.8.8)
Perl:DBI
Perl:DBD:mysql
Andy S-C/IBM perl MQTT framework.
Linux Bluez stack and tools

MQTT Broker
MySQL with Bluelocate Database (included) loaded

================================================
Install Instructions:

> Database Logger  /bluelocate
Create a new database on your MySQL server.
Create the database table using the contents of bluelocateDB/structure.sql
Copy the bluelocate directory to somewhere useful.
Copy the shared directory from the perl Framework to the same tree.
Edit mqtt.pl to include the bluelocate module, using the existing module
definitions as a template.
Setup the database handle in bluelocate/bluelocate.pl lines 18 and 26
Set the topic space in bluelocate/bluelocate.pl line 14
Set the broker details in shared/mqtt.pl

>Bluetooth scanner /bluemonitor
Copy the bluemonitor directory to somewhere useful.
Copy the shared directory from the perl Framework to the same tree.
Edit mqtt.pl to include the bluemonitor module, using the existing module
definitions as a template.
```

```
Set the broker details in shared/mqtt.pl
Set the topic space in bluemonitor/bluemonitor.pl line 14
Set hcitool location in bluemonitor/bluemonitor.pl line 27

============================
Execution Instructions:

> Database Logger  /bluelocate
Change to the bluelocate directory.
Execute ../shared/mqtt.pl bluelocate

> Bluetooth Scanner  /bluemonitor
Change to the bluemonitor directory.
Execute ../shared/mqtt.pl bluemonitor
```

B.2.8 CombeChem remote control tool

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

============================
Dependencies

LEJOS libraries
IBM MQTT Java Classes IA92
IBM Perl-CGI module (for web interface)
Linux drivers for camera and Lego IR Tower

============================
Directory Contents.

lego_mqtt/
   The RCX, IR-Bridge and camera control code.
 Usage instructions within.

lego_www/
   The web interface directory.
 To use, Copy content to Web Server
   Install the MQTT Perl-CGI
   Update index.html with the correct cgi location
   Update webcam.php and webcam2.php to point at the image grab on the
      webcam servers.
```

B.2.9 CombeChem remote control modules

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.
This work was inspired by java source code from Dave Conway-Jones and Andy
Stanford-Clarke at IBM's Hursley Laboratory.

============================
Dependencies

LEJOS libraries
IBM MQTT Java Classes IA92
Linux drivers for camera and Lego IR Tower

============================
Directory Contents.

intelscope/
   Java module to control the intelplay microscope through the linux
 /proc/ filesystem

RCX/
   Java modules for the Mindstorms RCX and IR-Bridge.

============================
Usage - intelscope

   Copy directory contents onto the machine the camera connects to.
   Check the drivers load ok (with a Video4Linux Client) and that the /proc/
      directory i created correctly.
   Allow the user running the code Read/Write permissions to the file
      /proc/cpia.
   Edit go.sh to include the correct proc entry, broker and broker clientID
   Run go.sh
```

```
============================
Usage - RCX

   Copy directory contents onto the machine the irbridge connects to.
   Install the lejos tools, following the supplied instructions.
   Reflash the RCX with the lejos Firmware, and LegoMicroscope.bin codebase,
      again following the lejos instructions.
   Reboot the RCX!
   On the PC edit go.sh, to include the correct variables at the top.
   Run go.sh
```

B.2.10 CombeChem MQTT/Database Enabled SHG capture

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

This code is based on, and includes redistributables from, IBM IA93 -
"WebSphere MQ Telemetry Transport C language implementation" and may contain
additional licensing restrictions.  Please refer to included IA93_Lic_en.tx

============================
Dependencies

Visual Basic 6
Microsoft Data Access Components
Postgres ODBC driver
vcomsoft UTIMEc Library
Dictionary2MD5 library
National Instruments Measurement Studio Components

============================
Install Instructions.

Contact Jeremy Frey (jgf@soton.ac.uk) as multiple parts of this software
have distribution restrictions.
```

B.2.11 CombeChem SHG Experiment AutoBlogger

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

============================
Dependencies

Perl (tested with version 5.8.8)
Perl:DBI
Perl:DBD:mysql
Perl:DBD:Pg
Andy S-C/IBM perl MQTT framework.
R (tested with version 2.4.0)
GNU Plot

MQTT Broker, receiving messages from either SHG Rig Control software, or
the experiment database web interface.

SHG Experiment Database
Raw Access to Chemtools Blog database.

============================

Install Instructions:

Copy the autoR directory to somewhere useful.
Copy the shared directory from the perl Framework to the same tree.  Edit
mqtt.pl to include the autoR module, using the existing module definitions
as a template.

Setup the database handlers in lines 15-30 of autoR/blogger.pl
Set the broker details in shared/mqtt.pl
Set the broker topic in autoR/autoR.pl line 19
Set GnuPlot exectuable in autoR/blogger.pl line 18
Set the ODBC source in Export_SHG.R line 8
```

```
============================
Execution Instructions:

Change to the autoR directory.
Execute ../shared/mqtt.pl autoR
```

B.2.12 CombeChem SHG Experiment database to review database, and automatic fit tools

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

============================
Dependencies

Perl (tested with version 5.8.8)
Perl:DBI
Perl:DBD:mysql
Perl:DBD:Pg
R (tested with version 2.4.0)
GNU Plot

SHG Experiment Database
Review data Database

============================
Install Instructions:

Copy the directories to somewhere useful.

>>Experiment extractor
Set GnuPlot exectuable in extract.pl line 16
Setup the database handlers in lines 18-26 of extract.pl
Set R exectuable in extract.pl line 135
Set ODBC source (experiment database) in Export_SHG.R line 8

>>Experiment fit tool
Setup the database handlers in lines 18-20 of fitall.pl
Set GnuPlot exectuable in autoR/blogger.pl line 18
Set R exectuable in extract.pl line 82
```

```
============================
Execution Instructions:

>>Experiment extractor
execute extract.pl
Go make Tea! (It takes a while on a big experiment database).
If it fails or is stopped, then it can be restarted and will continue
from where it stopped.

>>Experiment fit tool
execute fitall.pl
Go make Tea! (It takes a while on a big experiment database).
If it fails or is stopped, then it can be restarted and will continue
from where it stopped.
```

B.2.13 CombeChem SHG Web frontend to the experiment and review databases

```
(c) University of Southampton 2005
Author - Jamie Robinson, School of Chemistry.

Example code provided with no warranty, or guarantee, only to demonstrate
the work undertaken during the authors PhD.
Any use of this code is at your own risk, Jamie Robinson and the University
of Southampton cannot be held responsible for the consequences of any use.

============================================
Dependencies

Webserver (tested with Apache 2) with PHP (tested with Php4)
PHP_GNUPlot (included)
GNU Plot
SHG Experiment Database
Review data Database

============================================
Install Instructions:

Copy the directory to somewhere useful on your webserver.

>>Experiment database tool (smartSHG)
Set the ODBC source(experiment DB) in auth.php

>>Experiment review tool (expt_review)
Check that the variables at the top of PHP_GNUPlot.php are sane.
Set the ODBC source(experiment DB) in shgDB_tool.php
set the review database host/user/password at the top of all the php files.

============================================
Execution Instructions:

>Experiment database tool (smartSHG)
Access index.html via your web-browser.

>Experiment review tool (expt_review)
Access index.php via your web-browser.
```

Publication List

[1] Jamie M Robinson, Jeremy G Frey, Lefteris Danos, and Honchen Fu. Design of grid computing infrastructure to aid second harmonic generation studies at the liquid/air interface. In *Chemical Physics: Perspectives and Prospects for the Future*, UK, July 2004. Institute of Physics.

[2] Jamie M. Robinson, Jeremy G. Frey, Andy J. Stanford-Clark, Andrew D. Reynolds, and Bharat V. Bedi. Chemistry by mobile phone (or how to justify more time at the bar). In *UK e-Science All Hands Meeting 2005*, UK, September 2005.

[3] Esther R. Rousay, Hongchen Fu, Jamie M. Robinson, Jonathan W. Essex, and Jeremy G. Frey. Grid-based dynamic electronic publication: a case study using combined experiment and simulation studies of crown ethers at the air/water interface. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, 363(1833):2075–2095, August 2005.

[4] Jeremy G. Frey, Hugo R. Mills, Gareth V. Hughes, Jamie Robinson, Dave C. de Roure, monica c. schraefel, and Luc Moreau. Semantic support for smart laboratories. In *229th American Chemical Society (ACS) National Meeting*, USA, March 2005. American Chemical Society. CombeChem output.

[5] Jeremy G. Frey, Jamie Robinson, Andrew Stanford-Clark, Andrew Reynolds, and Bharat Bendi. *Sensor Networks and Grid Middleware for Laboratory Monitoring.* IEEE Computer Society, USA, December 2005. CombeChem Project Output.

[6] Jamie M. Robinson, Jeremy G. Frey, D. C. De Roure, Andrew J. Stanford-Clark, Andrew D. Reynolds, Bharat V. Bedi, and D. Conway-Jones. The combechem mqtt lego microscope: a grid enabled scientific apparatus demonstrator. In *Proceedings of the Fifth All Hands Meeting*, pages 393–396, UK, October 2006. National eScience Center.

[7] R. Greef, J.G. Frey, J. Robinson, and L. Danos. Adsorption of rhodamine 6g at the water-air interface. *Physica Status Soilidi (C)*, 5(5):1187–1189, June 2007.

[8] R. Greef, J.G. Frey, J. Robinson, and L. Danos. Adsorption of rhodamine 6g at the water-air interface (in special issue: 4th international conference on spectroscopic ellipsometry (icse4)). *physica status solidi (c)*, 5(5):1187–1189, May 2008.

References

[1] P. W. Atkins. *Physical Chemistry*, chapter 6, page 154. OUP, 6th edition, 1998.

[2] V. Fried, H. Hameka, and U. Blukis. *Physical Chemistry*, chapter 6, pages 140–141. Macmillan Publishing, 1st edition, 1977.

[3] W. J. Moore. *Physical Chemistry*, chapter 6, page 206. Longman, 5th edition, 1972.

[4] W. J. Moore. *Physical Chemistry*, chapter 11, pages 484–486. Longman, 5th edition, 1972.

[5] R. M. A. Azzam and N. M. Bashara. *Ellipsometry and Polarised Light*, chapter 3,6. North Holland Publishing Company, 1st edition, 1977.

[6] R. W. Boyd. *Nonlinear Optics*, chapter 1, pages 1–2. Academic Press, 1st edition, 1992.

[7] K. B. Eisenthal. Liquid interfaces probed by second-harmonic and sum-frequency spectroscopy. *Chem. Rev.*, 96(7):1343–1360, 1996.

[8] R. M. Corn and D. A. Higgins. Optical second harmonic generation as a probe of surface chemistry. *Chem. Rev.*, 94:107–125, 1994.

[9] P. B. Miranda and Y. R. Shen. Liquid interfaces: A study by sum-frequency vibrational spectroscopy. *J. Phys. Chem. B*, 103(17):3292–3307, April 1999.

[10] P. J. Campagnola and L. M. Loew. Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. *Nature Biotechnology*, 21:1356–1360, 2003.

[11] J. G. Frey. Comb-e-chem: an e-science research project. In *EuroQSAR 2002 Designing Drugs and Crop Protectants: processes, problems and solutions, Oxford, UK,*, pages 395–398. University of Southampton, Blackwell, 2002.

[12] CLEF. Clinical e-Science Framework. http://www.clinical-escience.org/. Accessed 24 September 2007.

[13] BiosimGRID. A GRID Database for biomolecular simulations. http://www.biosimgrid.org/. Accessed 24 September 2007.

[14] DAME. Distributed Aircraft Maintenance Environment. http://www.cs.york.ac.uk/dame. Accessed 24 September 2007.

[15] CoAKTinG. Collaborative Advanced Knowledge Technologies in the grid. http://www.aktors.org/coakting. Accessed 24 September 2007.

[16] MYGRID. Directly Supporting the E-Scientist. http://www.mygrid.org.uk/. Accessed 24 September 2007.

[17] CombEchem. Structure-Property Mapping: Combination Chemistry and the Grid. http://www.combechem.org/. Accessed 24 September 2007.

[18] schraefel mc, Hughes G., Mills H., Smith G., and Frey J. Making tea: Iterative design through analogy. In *Proceedings of Designing Interactive Systems*, 2004.

[19] S.J. Coles, J.G. Frey, M.B. Hursthouse, M.E. Light, A.J. Milsted, L.A. Carr, D. DeRoure, C.J. Gutteridge, H.R. Mills, K.E. Meacham, M. Surridge, E. Lyon, R. Heery, M. Duke, and M. Day. An e-science environment for service crystallography-from submission to dissemination. *Journal of Chemical Information and Modeling*, 46(3):1006–1016, 2006.

[20] E. R. Rousay, H. Fu, J. M. Robinson, J. W. Essex, and J. G. Frey. Grid-based dynamic electronic publication: a case study using combined experiment and simulation studies of crown ethers at the air/water interface. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, 363(1833):2075–2095, August 2005.

[21] R.A. Mansson, J.G. Frey, J.W. Essex, and A.H. Welsh. Prediction of properties from simulations: a re-examination with modern statistical methods. *Journal of Chemical Information and Modeling*, 45(6):1791–1803, December 2005.

[22] R. A. Mansson, A. H. Welsh, J. G. Frey, and L. Danos. Designing experiments for an application in laser and surface chemistry. May 2003.

[23] M. Hazas, J. Scott, and J. Krumm. Location-aware computing comes of age. *Computer*, 37(2):95–97, February 2004.

[24] N. G. Leveson. *Safeware: System Safety and Computers*, chapter Medical Devices: The Therac-25. 1995. Addison-Wesley. Also available online at http://sunnyday.mit.edu/papers/therac.pdf.

[25] N. G. Leveson and C. S. Turner. An investigation of the therac-25 accidents. *Computer*, 26(7):18–41, July 1993.

[26] Gareth Hughes, mc schraefelel, and Jermey G Frey. Breaking the book: Translating the chemistry lab book into a pervasive computing lab environment. 2003.

[27] A. Dix. Absolutely crackers, designing an experience. http://www.hcibook.com/alan/papers/crackers2001/, November 2001.

[28] CambridgeSoft Corporation. Chemoffice. http://www.camsoft.com. Accessed 24 September 2007.

[29] NoteBookMaker ltd. Notebookmaker. http://www.notebookmaker.com/. Accessed 24 September 2007.

[30] Scrip-Safe. http://www.scrip-safe.com/laboratory_notebooks.htm. Accessed 24 September 2007.

[31] Intellichem Ltd. Suppliers of Electronic Lan Notebook Solutions. http://www.intellichem.com, 2004. Accessed 24 September 2007.

[32] L. Arnstein, G. Borriello, S. Consolvo, C. Hung, and J. Labscape Su. A smart environment for the cell biology laboratory. *IEEE Pervasive Computing Magazine*, 1(3):13–21, July-September 2002.

[33] W. E. Mackay, G. Pothier, C. Letondal, K. Boegh, and H. E. Sorensen. The missing link: augmenting biology laboratory notebooks. In *Proceedings of the 15th annual ACM symposium on User interface software and technology*, pages 41–50. ACM Press, 2002.

[34] J. D. Myers. Collaborative electronic notebooks as an electronic records: Design issues for the secure electronic laboratory notebook (eln). In *Procedings of the International Symposium on Collaborative Technologies and Systems*, 2003.

[35] J. G. Frey and M. B. Hursthouse. From e-science to publication@source. In *National Policies on Open Access (OA) Provision for University Research Output, Southampton, UK*. University of Southampton, 2004.

[36] CCDC. The cambridge structural database - the world repository of small molecule crystal structures. http://www.ccdc.org.uk/products/csd/. Accessed 24 September 2007.

[37] RealVNC. http://www.realvnc.com/. Accessed 24 September 2008.

[38] A. Stanford-Clark. Integrating monitoring and telemetry devices as part of enterprise information resources. Technical report, IBM, 2002.

[39] IBM. Mqtt.org. http://www.mqtt.org. Accessed 24 September 2007.

[40] IBM. WebSphere MQ Telemetry Transport Specification. http://publib.boulder.ibm.com/infocenter/wbihelp/index.jsp?topic=/com.ibm.etools.mft.doc/ac10840_.htm. Accessed 24 September 2007.

[41] A. Stanford-Clark. Private correspondance, 2004-2007.

[42] IBM. WebSphere Business Integration Message Broker. http://www-306.ibm.com/software/integration/wbimessagebroker. Accessed 24 September 2007.

[43] FloodNet. Pervasive Computing in the Environmment. http://envisense.org/floodnet/floodnet.htm. Accessed 24 September 2008.

[44] Singleton Labs. MonarchCharts - Java. http://www.singleton-labs.com/mcharts.php. Accessed 24 September 2007.

[45] IBM. IA93: WBI Brokers - C implementation of WebSphere MQ Telemetry transport. http://www-1.ibm.com/support/docview.wss?rs=171&uid=swg24006525&loc=en_US&cs=utf-8&lang=en. Accessed 24 September 2007.

[46] IBM. IA92: WBI Brokers - Java implementation of WebSphere MQ Telemetry transport. http://www-1.ibm.com/support/docview.wss?rs=171&uid=swg24006006&loc=en_US&cs=utf-8&lang=en. Accessed 24 September 2007.

[47] University of Southampton. Information systems services regulations. http://www.soton.ac.uk/iss/regs/index.html. Accessed May 2008.

[48] S. J. Wilson. Middleware infrastructure for remote experiment monitoring and control. Master's thesis, School of Chemsitry, University of Southampton, 2007.

[49] M. Nixon. *Image Processing and Feature Extraction*. Newnes, 2002.

[50] Highways Agency. M42 jct 3a - jct 7 active traffic management. http://www.highways.gov.uk/roads/projects/4673.aspx. Accessed 15 September 2007.

[51] Transport for London. Tfl - congestion charging homepage. http://www.cclondon.com/. Accessed 15 September 2007.

[52] O. Kwon, S. Choi, and G. Park. Nama: a context-aware multi-agent based web service approach to proactive need identification for personalized reminder systems. *Expert Systems with Applications*, 29(1):17–32, 2005.

[53] IBM Corporation. IBM Student Researchers Develop Mobile Phone Technology to Help Hearing Impaired. http://www.ibm.com/news/se/sv/2006/09/12-forskare.html.

[54] F.J. Gonzalez-Castano, J. Garcia-Reinoso, F Gil-Castineira, E Costa-Montenegro, and J.M. Pousada-Carballo. Bluetooth-assisted context-awareness in educational data networks. *Computers and Education*, 45:105–121, 2005.

[55] Digium Inc. Asterisk :: The open source pbx & telephone platform. http://www.asterisk.org/. Accessed May 2008.

[56] voip info.org. Asterisk bluetooth channels. http://www.voip-info.org/wiki/index.php?page=Asterisk+Bluetooth+channels. Accessed May 2008.

[57] J Evans, D Hemment, T Humphries, and M Raento. Loca: Location oriented critical arts. http://leoalmanac.org/gallery/locative/loca/index.htm, 2006.

[58] A. J. Milsted. Facilitating chemical discovery: An e-science approach. Master's thesis, School of Chemsitry, University of Southampton, 2007.

[59] Internet2 Middleware Initiative. Shibboleth. http://shibboleth.internet2.edu/. Accessed August 2008.

[60] Lejos : Java for the rcx. http://lejos.sourceforge.net/. Accessed 24 April 2006.

[61] Cpia webcam driver for linux. http://webcam.sourceforge.net/. Accessed 24 April 2006.

[62] A. Nirendra. Exploring procfs. *Linux Gazette*, 115, June 2005. http://linuxgazette.net/115/nirendra.html, Accessed 20 September 2005.

[63] jtravis@p00p.org. Official camserv home page. http://cserv.sourceforge.net. Accessed 28 April 2006.

[64] P. A. Franken, A. E. Hill, C. W. Peters, and G. Weinreich. Generation of optical harmonics. *Phys. Rev. Lett.*, 7(4):118–119, Aug 1961.

[65] I. Freund and P. M. Rentzepis. Second harmonic generation in liquid crystals. *Phys. Rev. Lett.*, 18(11):393–394, February 1967.

[66] L. Danos. *Non-Linear Spectroscopic Studies at Interfaces: Experiment and Theory.* PhD thesis, Department of Chemsitry, University of Southampton, September 2003.

[67] M. J. Crawford. *Second Harmonic Generation From Liquid Interfaces.* PhD thesis, Department of Chemsitry, University of Southampton, January 1995.

[68] V. Mizrahi and J. E. Sipe. Phenomenological treatment of surface second-harmonic generation. *J. Opt. Soc. Am. B*, 5(3):660–667, March 1988.

[69] D. A. Higgins, S. K. Byerly, M. B. Abrams, and R.M. Corn. Second harmonic generation studies of methylene blue orientation at silica surfaces. *J. phys. Chem.*, 95(18):6984–6990, 1991.

[70] P. F. Brevet. Phenomenological three-layer model for surface second-harmonic generation at the interface between two centrosymmetric media. *J. Chem. Soc., Faraday Trans.*, 92:4547–4554, 1996.

[71] A. J. Fordyce, W. J. Bullock, A. J. Timson, S. Haslam, R. D. Spencer-Smith, A Alexander, and J. G. Frey. The temperature dependence of surface second-harmonic generation from the air-water interface. *Molecular Physics*, 99:677–687, 2001.

[72] J. Bleecker. A manifesto for networked objects – cohabiting with pigeons, arphids and aibos in the internet of things. *Near Future Laboratory*, February 2006.

[73] A. H. Welsh, R. A. Manson, Frey J. G., and L. Danos. Statistical analysis of second harmonic generation experiments: a phenomenological model. *Chemometrics and intelligent laboratory systems*, 75:45–54, 2005.

[74] F. L. Arbeloa, I. L. Gonzalez, Ojeda P. R., and Arbeloa I L. Aggregate formation of rhodamine 6g in aqueous solution. *J. Chem. Soc. Faraday Trans. 2*, 78:989–994, 1982.

[75] Ojeda P. R., F. L. Arbeloa, and Arbeloa I L. Dimerization and trimerization of rhodamine 6g in aqueous solution. *J. Chem. Soc. Faraday Trans. 2*, 84:1903–1912, 1988.

[76] D. Toptygin, B.Z. Packard, and L Brand. Resolution of adsorption spectra of rhodamin 6g addregates in aqueous solution using the law of mass action. *Chem. Phys. Lett*, 277:430–435, 1997.

[77] J. Lindsey. Photochemcad. http://www.photochemcad.com/.

[78] A. Fischer, C. Cremer, and E. H. K. Stelzer. Fluorescence of coumarins and xanthenes after two-photon adsorption with a pulsed titanium-sapphire laser. *Applied Optics*, 34:1989–2003, 1995.

[79] A. N. Adhikesavlu, D. Mastropalolo, A. Camerman, and N. Camerman. 9-(2-(ethoxycarbonyl)phenyl)-3,6-bis(ethylamino)-2,7-dimethylxanthylium chloride monohydrate. *Acta Crystallogr., Sect. C:Cryst. Struct. Commun.*, 57:657, 2001.

[80] K. Seno, T. Ishioka, A. Harata, and Y. Hatano. Photoionization of rhodamine dyes adsorbed at the aqueous solution surfaces investigated by synchrotron radiation. *Analytical Sciences 2001*, 17:i1177–i1179, 2001.

[81] R. Greef, J.G. Frey, J. Robinson, and L. Danos. Adsorption or rhodamine 6g at the water-air interface. *Phsica Status Solidi C*, 5(5):1187 – 1189, June 2007.

[82] WikiPedia. Water. http://en.wikipedia.org/wiki/Water.

[83] L Hecht and LD Barron. New aspects of second-harmonic optical activity from chiral surfaces and interfaces. *Molecular Physics*, 89(1):61–80, September 1996.

[84] A. Rajasekar, M. Wan, R. Moore, W. Schroeder, G. Kremenek, A. Jagatheesan, C. Cowart, B. Zhu, SY. Chen, and R. Olschanowsky. Storage resource broker - managing distributed data in a grid. *Computer Society of India Journal, Special Issue on SAN*, 33(4):42–54, October 2003.