

**Chemical Applications of *e*Science to
Interfacial Spectroscopy.
MPhil/PHD Transfer Report**

Jamie Robinson

September 2004

School of Chemistry,
University of Southampton,
j.m.robinson@soton.ac.uk

Supervisor Dr J.G. Frey
Advisor Prof. J. Dyke

Contents

1	Introduction	1
2	Interfacial Chemistry Background	3
2.1	Why study liquid interfaces?	3
2.2	Liquid Surface probing techniques.	3
2.2.1	Surface Tension.	3
2.2.2	Non-Linear Spectroscopy (SFG and SHG)	3
2.3	Second Harmonic Generation	4
2.3.1	Basic theory of SHG.	4
2.3.2	Nature of Second Harmonic Generated Emissions.	5
2.3.3	Origins of the Liquid Surface Specificity of SHG	6
2.3.4	Intensity of the Second Harmonic Signal.	6
2.3.5	The effect of polarisation on Second Harmonic Signal.	8
2.3.6	Why use SHG as an example	9
3	eScience Background	11
3.1	What is this whole eScience thing?	11
3.2	Combinatorial Approach - <i>More Chemistry Quicker</i>	11
3.3	“SmartLabs” and Data Loggers	12
3.4	Real-time Data Communication.	13
3.4.1	Publish/Subscribe and Broker Technologies.	14
3.5	Data Storage and Accessibility	15
3.6	Publish@Source, Publish in Full.	18
4	Experimental	19
4.1	Data System Overview	19
4.2	SmartLab Environment Data	20
4.2.1	Electronics	20

4.2.2	Publish/Subscribe model data broker.	21
4.2.3	Storage	22
4.2.4	Data Recall	23
4.3	SHG Experimental	26
4.3.1	Optics and Electronics	26
4.3.2	Sample Studied	27
4.3.3	Software	28
4.3.4	Managing system inconsistencies	29
4.3.5	Analysis (Current) of SHG data.	31
5	Results	33
5.1	“System Inconsistency Management” data	33
5.1.1	Power Stability	33
5.1.2	Signal Stability	33
5.1.3	Monochromator Scan	35
5.2	Raw data	37
6	Future Work	39
7	Glossary	41
7.1	Chemistry Terms	41
7.2	Computer Science Terms	46
	References	52

List of Figures

1	SmartLab Data Flow Overview.	19
2	Circuit for the monitoring of toggle sensors.	21
3	Sample Message for the SmartLab Broker system.	22
4	Data Flow from Sensor to Client	23
5	Example Screen-Shot of the lab monitoring webpage.	24
6	Screen-shot showing part of the "labrecall" page	25
7	Optics of the SHG Experiment	26
8	Relational Database Scheme for storage of SHG related data.	28
9	Analysis work-flow for SHG polarisation dependence	32
10	Power Stability of the SHG lab MiniLite Laser.	33
11	Scan to test the stability of a weak R6G signal.	34
12	Stability plot for a 5×10^{-6} M R6G solution.	34
13	Monochromator Scan showing the strong fluorescence peak observed in this sample.	36
14	A more ideal monochromator scan, showing good separation of the SHG peak from the background.	36
15	Polarisation dependence scan for 1×10^{-7} M R6G in water	37
16	Polarisation dependence scan for 1×10^{-6} M R6G in water	38
17	Polarisation dependence scan for 1×10^{-5} M R6G in water	38
18	Schematic of light reflecting and refracting from an interface between materials of refractive index n_1 and n_2 . [6]	43

Acknowledgements

Engineering and Physical Sciences Research Council

Combechem Researchers

Frey Group

Andy Standford-Clark, Andrew Reynolds and Colleagues at IBM Hursley

Dr Jeremy G Frey

Dr Lefteris Danos

Foreword

This report is a summary of works carried out by the author between October 2003 and September 2004, in the first year of his PhD.

Initial work involving SHG was carried out under the supervision of Dr Lefteris Danos, continuing (and drawing upon) his work involving Rhodamine 6G (R6G), it is hoped that the work with R6G will be concluded and published shortly. Initial work on Broker technologies was carried out with assistance from Dr Gareth Hughes in the School of Electronics and Computer Science.

Due to the multi-disciplinary nature of the research being carried out, the readership will include Computer Scientists as well as Chemists. To aid this a Glossary is provided in chapter 7. This is split into two sections, one explaining the more common chemistry terms used, and one the computer science terms.

Attached is a CD, containing the software written to date. Much of this relies on the availability of network servers, and isn't designed to run standalone. Also included on the CD are electronic (high resolution) PDF versions of the images in this report, and a pdf version of the report.

It should be remembered that this research is a work in progress, and at this stage it is difficult to draw many conclusions. Some raw results are presented in chapter 5, and an outline of future work is given in chapter 6.

This document was typeset in 12 point palatino, using LyX, TeX and Debian Linux.

1 Introduction

Modern synthetic chemistry is fast heading towards an explosion in synthetic production. Combinatorial chemistry techniques are providing the methods to produce vast numbers of similar, but distinct, molecules. Production of these molecules isn't the end of the story, they then need to be analysed, characterised and logged. As much of this work is target towards medicinal science, one also needs to produce a full audit trail to prove the provenance of the products. Tradition paper based recording methods can no longer cope with the vast amounts of data. A new solution is needed to deal with data from these types of experiment, from chemists conception through to published paper. Work is increasingly been done in a collaborative fashion, in the traditional work methodology this is generally limited to groups within the same building - the paper postal service simply takes too long! By digitising and automating systems these physical boundaries can be overcome using Internet technologies, whilst still maintaining security and access control over who can access your data. Once data is available in a digital medium, publishing it becomes a more trivial matter. By starting the digital storage procedures from the ground up, one can also build in provenance storage and reproduction. By publishing this provenance and raw data alongside the final analysed results the reader is empowered to make their own decisions about reliability and authenticity. In this way a publish@source ideal can be realised where the current unpublished wealth of information can be made available. This can occur without necessarily requiring the data to be fully peer-reviewed, therefore removing the political bottle neck to sharing of information. There is of course the potential for unreliable data to make it's way into the public domain This data would traditionally have been reduced by firstly publisher pride, and secondly by the peer-review process. However the volume of potential data that could be published should be able, at the very least, to cast doubt on results through conflicting data sets and, hopefully, thought significant numbers of corresponding data sets one may be able to draw confident conclusions that a lone researchers work wouldn't allow.

Liquid surfaces play a fundamental role throughout Science and Technology, embracing fields as widely spaced as biological systems and battery technologies. It is therefore important to develop a range of techniques that can characterise the behaviour of these interfaces, it is reasonably obvious that they will

differ greatly from the bulk materials. The relatively small system size (often only the first few molecules depth at the surface) makes this an interesting challenge, however the very nature of the surface can make make this somewhat easier. Liquid bulk is by definition isotropic, implying that a centre of inversion exists. At the surface this symmetry is broken, and this asymmetry can be exploited by some light based techniques. Second Harmonic Generation (SHG) is an example of one such experiment, SHG is electric dipole forbidden in centrosymmetric media[1], hence for liquid-liquid and liquid-gas interfaces is inherently surface specific. It involves (in the experiment reported here) reflecting a laser beam off a liquid-air or liquid-liquid surface, and monitoring radiation released at half the incident wavelength - the Second Harmonic radiation. By monitoring the output radiation intensity and polarisation with respect to the polarisation of the incident radiation, it is possible to deduce information about the alignment of SHG active components at the interface[1]. To simplify collection of this data, the experiment has received some computer experiment control in the past, including automated data collection and automated control of input polarisation. Once the data is collected, the use of computers becomes some what more ad-hoc with individual scientists having the own recipe of scripts, ranging from desktop spreadsheets to statistical analysis packages. This initial work has resulted in a reliable digital data stream, that can be wrapped and packaged into an *eScience* data scheme. To aid this, techniques such as databasing, automation and meta-data enforcement can improve and extend the data stored, and hence ease recall, analysis and publishing. By extending the computer controlled automation of the experiment it will be possible to improve throughput, improve data quality and release the experimenter to other tasks.

2 Interfacial Chemistry Background

2.1 Why study liquid interfaces?

Liquid interfaces play a fundamental role in science and technology. In particular the air/liquid and liquid/liquid interfaces have a fundamental role in biological systems, environmental processes and surfactant chemistry to name a few. 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. Over the past decades, a range of techniques have been developed to support this.

2.2 Liquid Surface probing techniques.

2.2.1 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 surface tension of a liquid, γ , is defined as the amount of work, δw , required to change the surface area by some amount, $\delta\sigma$. This can be formalised as $\delta w = \gamma \cdot \delta\sigma$ [2]. The surface tensions of many liquids are well defined, and deviations from these values can be an indication of impurities. However Surface Tension is a macroscopic measurement that yields no information about the nature of the adsorbed species.

2.2.2 Non-Linear Spectroscopy (SFG and SHG)

Non-linear optical techniques are inherently surface specific in fluid systems, non-linear effects are electric dipole forbidden in centrosymmetric media. Two important non-linear techniques are those of second harmonic generation and sum frequency generation[3, 4, 5].

In the Sum Frequency Generation experiment a sample is subjected to two light sources of differing frequencies (ω_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 .

From the electric dipole approximation, it follows that SHG only occurs in non-centrosymmetric media[3], therefore it is a useful probe that can investigate 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 specificity also allows for monitoring of structure and properties down to sub-monolayer coverage levels.

2.3 Second Harmonic Generation

This section assumes basic familiarity with the electromagnetic nature of light (15 on page 45), polarisation (17 on page 45) and principle of refractive index (18 on page 46), as described in the glossary.

2.3.1 Basic theory of SHG.

When a light wave is incident on a dielectric material, its electric field causes a distortion to the charge distribution within that material. This distortion causes an induced dipole moment, which re-radiates at the same frequency (due to conservation of energy laws). The sum of the dipole moments per unit volume is called the electric polarisation \mathbf{P} , and is proportional to the electric field vector \mathbf{E} , hence

$$\mathbf{P} \propto \mathbf{E} \quad (1)$$

If we assume the 1 dimensional case, \mathbf{P} and \mathbf{E} become scalar. The proportionality constant in this case is $\chi^{(1)}$, the electric susceptibility of the medium, multiplied by ϵ_0 , the permittivity of free space[6], giving

$$P = \epsilon_0 \chi^{(1)} E \quad (2)$$

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:

$$P = \epsilon_0(\chi^{(1)}E + \chi^{(2)}E^2 + \chi^{(3)}E^3 + \dots) \quad (3)$$

where $\chi^{(2)}$ and $\chi^{(3)}$ are the second and third order electric susceptibilities of the medium. These give rise to the generation of doubled and tripled frequency radiation. If we describe the electric field associated with a light wave as a cosine function dependant on frequency, we could write:

$$E = E_0 \cos(\omega t) \quad (4)$$

Hence when this light wave is incident on a dielectric medium, combining 3 and 4 gives the resulting polarisation:

$$P = \epsilon_0(\chi^{(1)}E_0 \cos(\omega t) + \chi^{(2)}E_0^2 \cos^2(\omega t) + \chi^{(3)}E_0^3 \cos^3(\omega t) + \dots) \quad (5)$$

Then considering the second order polarisation term:

$$P^{(2)} = \epsilon_0 \chi^{(2)} E_0^2 \cos^2(\omega t) \quad (6)$$

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

$$P^{(2)} = \frac{1}{2} \epsilon_0 \chi^{(2)} E_0^2 [\cos(2\omega t) + 1] \quad (7)$$

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.

2.3.2 Nature of Second Harmonic Generated Emissions.

We can also see from equation 6 that the intensity of the second harmonic emission is dependant 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, 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. 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[7].

2.3.3 Origins of the Liquid Surface Specificity of SHG

Equation 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 opposite direction polarisation of $-P_\omega$. If we now consider second harmonic generation case, equation 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 we would expect the opposite electric field $-E_\omega$ to give polarisation in the opposite direction, $-P_{2\omega}$. However equation 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, hence 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).

2.3.4 Intensity of the Second Harmonic Signal.

In their 1988 paper Mizrahi and Sipe[8] derived the following equation linking SHG signal to experimental parameters for the reflection configuration, as used for this work.

$$I_{2\omega} = \frac{32\pi^3\omega^2}{c^3A} \sec^2\theta_{in} |\hat{\mathbf{e}}_{2\omega} \cdot \chi^{(2)} : \hat{\mathbf{e}}_\omega \cdot \hat{\mathbf{e}}_\omega|^2 I_\omega^2 \quad (8)$$

where the subscripts ω and 2ω indicate parameters calculated for the fundamental and a second harmonic 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{e}_ω 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{e}_ω can be broken down into two components:

$$\hat{e}_\omega = \bar{e}_\omega \cdot e_\omega \quad (9)$$

where e_ω is the Fresnel coefficients that relates the incident electric field to that in the interface, and \bar{e}_ω the unit vector describing the polarisation of the incident radiation field.

It is convenient to describe the vector \hat{e}_ω 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 lies parallel to the Y axis, and the the p polarised beam lies in the XZ plane.

Therefore we can express the vector \bar{e}_ω in equation 9 above, in terms of s and p polarisations, giving:

$$\bar{e}_\omega = \begin{pmatrix} s_\omega \\ p_\omega \end{pmatrix} = \begin{pmatrix} \sin \gamma \\ \cos \gamma \end{pmatrix} \quad (10)$$

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

Hence by combining this with e_ω in component form, equation 9 can be expressed

$$\hat{e}_\omega = \begin{pmatrix} E_{X\omega} \\ E_{Y\omega} \\ E_{Z\omega} \end{pmatrix} = \begin{pmatrix} \cos \gamma e_{X\omega} \\ -\sin \gamma e_{Y\omega} \\ \cos \gamma e_{Z\omega} \end{pmatrix} \quad (11)$$

The term $\chi^{(2)} : \hat{e}_\omega \cdot \hat{e}_\omega$ in equation 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[6, 7] that by expanding $\chi^{(2)}$, a third rank tensor, the following holds

$$\mathbf{P}^{\text{NL}} = \begin{bmatrix} 0 & 0 & 0 & \chi_{XYZ} & \chi_{XXZ} & 0 \\ 0 & 0 & 0 & \chi_{XXZ} & -\chi_{XYZ} & 0 \\ \chi_{ZXX} & \chi_{ZXX} & \chi_{ZZZ} & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} E_{X\omega}^2 \\ E_{Y\omega}^2 \\ E_{Z\omega}^2 \\ 2E_{Y\omega}E_{Z\omega} \\ 2E_{X\omega}E_{Z\omega} \\ 2E_{X\omega}E_{Y\omega} \end{bmatrix} \quad (12)$$

2.3.5 The effect of polarisation on Second Harmonic Signal.

By combining equations 8, 11 and 12 the following equations have been derived that relate to s out and p out polarisation of SHG[4, 6]

$$I_P^{(2\omega)} \propto |A \cos^2 \gamma + B \sin^2 \gamma|^2 I_\omega^2 \quad (13)$$

$$I_s^{(2\omega)} \propto |C \sin 2\gamma|^2 I_\omega^2 \quad (14)$$

$$A = a_2\chi_{XXZ} + a_3\chi_{ZXX} + a_4\chi_{ZZZ} \quad (15)$$

$$B = a_5\chi_{ZXX} \quad (16)$$

$$C = a_1\chi_{XXZ} \quad (17)$$

where the coefficients a_1 thru a_5 are functions of the refractive indices, the angles on incidence, reflection and transmittance, and the electric field components at the surface.

From the equations for the s and p output polarisation of SHG (equations 13 and 14) the equations for output polarisation of 45° and -45° can be derived :

$$I_{(+45)}^{(2\omega)} \propto \left| \frac{1}{\sqrt{2}} (A \cos^2 \gamma + B \sin^2 \gamma + C \sin^2 \gamma) \right|^2 I_\omega^2 \quad (18)$$

$$I_{(-45)}^{(2\omega)} \propto \left| \frac{1}{\sqrt{2}}(A \cos^2 \gamma + B \sin^2 \gamma - C \sin^2 \gamma) \right|^2 I^2 \quad (19)$$

By collecting data for S, P, and 45° output polarisations, then fitting to equations 13, 14 and 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 15, 16 and 17 :

$$\chi_{XXZ} = \frac{C}{a_1} \quad (20)$$

$$\chi_{ZXX} = \frac{B}{a_5} \quad (21)$$

$$\chi_{ZZZ} = \frac{A - \left(\frac{a_x C}{a_1}\right) - \left(\frac{a_3 B}{a_5}\right)}{a_4} \quad (22)$$

These values can then be related to the energy minimised molecular structures, and information about molecular orientation at the interface can be resolved[6].

2.3.6 Why use SHG as an example

One of the aims of the combechem project is to provide an end-to-end publishing mechanism, taking a molecule from synthesis to characterisation. One important step in this chain is characterisation of the substance. This could be covered by implementing a simple database storage for a technique such as NMR or FTIR where the results are reliable. However another goal of combechem 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 addition 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. Making additions such as these are somewhat more difficult when using commercial systems. The SHG experiment here uses stock electronics, and home-built control software, hence modifications to facilitate "SmartLab" techniques are somewhat easier to perform.

The analysis of SHG data is a reasonably involved procedure, during which a number of assumptions have to be made. By changing these assumptions, (slightly) different results may be obtained, hence it is important to record the exact steps performed in making that analysis. This fits in well with the publish@source ideas that combechem sets out to achieve.

3 eScience Background

3.1 What is this whole eScience thing?

“e-Science refers to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet.”[9]

“e-Science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it.”[10]

“[The Grid] intends to make access to computing power, scientific data repositories and experimental facilities as easy as the Web makes access to information. *Tony Blair, 2002*”[10]

Broadly speaking eScience is the use of computers to enhance science. The UK eScience project covers a vast range of topics including Clinical Science support (CLEF)[11], Biomolecular Simulation (BiosimGRID)[12], Distributed Aircraft Maintenance (DAME)[13], Collaborative Knowledge (CoAKTing)[14] and providing support methodologies such as storage and resource management (MyGRID)[15]. The work presented here has been within the Combechem[16] project investigating techniques for improving both the quality of data, and the throughput of the Non-Linear surface optical technique of Second Harmonic Generation. At a broader level, Combechem is investigating the chemical aspects of End-to-End data storage for the chemical process. This ranges from Electronic Lab Books in the Organic synthesis lab through automated crystal structure collection and resolution, Data collection for laser driven surface spectroscopy to ab-initio chemical property calculation, and the use of statistical methods to improve laboratory efficiency.

3.2 Combinatorial Approach - *More Chemistry Quicker.*

In chemical science there is an increasing trend toward high throughput, 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, and also in the increasing pervasiveness of instrumentation. Instruments that were previously expensive, and required specialist operators, are now be-

coming available in off the shelf low maintenance systems. Equally other experimental techniques are coming in alongside giving a wider range of potential data collection.

This ever increasing range of compounds and results presents a challenge to the scientific community as the traditional method of peer review publishing has already exceeded it's capacity in that large amount of chemical information remain unpublished within the academic community often forgotten or mislaid when it doesn't quite fit with the latest hot topic. By not publishing, this information become effectively lost, inaccessible to people outside the initial institution, and often forgotten by the scientists within it. By storing this information as it's collected into a well indexed system the results instantly becomes more useful, and then when required, retrieval becomes a far easier task. Indeed it would even be possible for other scientists within the institution, and beyond to find the information, and then gain access to it - subject to having adequate permissions of course.

There is however 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 alongside the final results, 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 wouldn't and conversely if contradicting results are shown, then confidence in the data may be reduced, which could prompt further experiments to confirm the correct conclusion and save making a incorrect conclusions prematurely.

3.3 "SmartLabs" and Data Loggers

As data handling techniques improve, larger quantities of data can be managed sensibly. Considering the improvements in compound production already mentioned, it is logical to conclude that spectroscopers will look for ways to improve on machine throughput. 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 better results as random "human" errors are removed (or at least significantly reduced). However we must remember that these systems are not infallible, and results that are not up to expectations should be questioned. It is likely that throughput may reach a level where there is no longer a human influence checking individual spectra. In these cases the computer systems should be equipped to spot errors. This can be in the form of monitoring the laboratory and apparatus environment, as inconsistencies here may lead to unreliable data, by recording these a human researcher can look back and check for problem signatures. If a well understood theory can lead to an in-silico prediction, then this could be checked against the recorded data, highlighting problem data for human checking. In a truly automated system however, one would want a computer agent to be doing this spotting, and re-scheduling samples for analysis where the results are substandard.

It has already been discussed that human factors can lead to anomalous data. One also needs to consider environmental factors. Some experimental techniques can be very sensitive to effects such as temperature, humidity, ambient light level, or vibration. It is important to keep a record of such conditions alongside experimental data. It is becoming more common to install systems in laboratories to control these factors, such as air conditioners, the danger that come with this is for one to assume that these systems function flawlessly. In this case it becomes reasonable to configure the environmental monitors to run independently from the experiment. However one then needs to be able to observe correlation between experiment and environment. If the laboratory is configured to run from a single time-source, then by recording a time-stamp alongside each piece of data, one can easily correlate multiple data sources.

3.4 Real-time Data Communication.

As experiments become more autonomous, the scientist is freed to perform other tasks during data collection. However it is important to periodically check the status of data capture. The most logical 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 lab may affect the results for example. In this case the ability to remotely view the experiment is useful. It may be adequate to use a remote screen viewing application such as VNC[17]. VNC allow machines

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, there is a maximum of 1 client. Equally there is often little need to provide remote control. 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 the 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 you also gain some attractive security bonuses. This data could even be displayed in a web-page (most web-server scripting languages can interface directly with databases) giving the ultimate in portability, well written web pages can be viewed by the vast majority of current systems, potentially from anywhere in the world.

Automated systems will often periodically need human intervention. In order to maximise productivity the system needs to be able to recognise when this interaction is needed, and summon appropriate assistance. There currently exist a wide range of options for contacting people, ranging from dedicated paging systems to reuse of consumer items like mobile phones. Possible options include sending emails, using instant messenger systems (eg IRC or MSN), Playing pre-recorded messages over the phone, or sending SMS/pager messages. A combination of these tools will be investigated to find suitable, cost effective messaging solutions.

3.4.1 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 being 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

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's relevant to them. Clients can be publishers of, and/or subscribers to, data and can range from big enterprise-based servers to hand-held pervasive computing devices, such as Palm handhelds, or unattended remote telemetry devices.

Publishers send units of information (called messages) to the broker, on a specific topic. The topic is like the subject line of an e-mail. It tells you what the body of the message contains. Subscribers register their interests in certain topics with the broker.

The broker manages connections from publishers and subscribers, and deals with authentication and access control lists, to control who is allowed to publish and subscribe to which topics. This helps ensure that an accounts department, for example, can't accidentally shut down an oil pipeline. The broker also transforms messages into a more useful format, like an XML document, and can apply business logic to route messages based on their topics or contents.[18]

It is important to remember that publishers and subscribers are usually unknown to each other, the broker acts as a matchmaker. This decoupling makes the system future-proof, because it generalises the use of data; it is not tied to specific applications. At any time, you can deploy new applications that can use a new combination of topics. It is this potential that makes publish/subscribe such a powerful concept.

Topics are arranged hierarchically, with slashes (/) between the levels, similar to a web URL. The hierarchy, which defines an information space, must be carefully designed to help ensure that data is available in a sensible, logical, structure.

3.5 Data Storage and Accessibility

To support the quantity of data that current techniques can produce, efficient storage techniques must be found in order for data loss to 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 a personal identification system. By providing centralised data resources these spectra can be stored in a reliable, retrievable manner.

But why stop at spectra? By extending the data storage systems to include the Lab book as well, an end to end data management systems 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's common (and useful) to have the lab book alongside you whilst performing preparations. However, something as simple as a solvent spillage could render the contents of the book useless. In an electronic system the same solvent spillage may render the terminal in useless, but all your data would be stored safely on the centralised server, ready to be recalled via another terminal. It is this security of data that has impressed practical chemists most with these types of scheme. A recent study produced the quote "I can go anywhere and it's, like, this is me and my data. It's all there! Bang!"[19] which related a chemist's 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"[20]. This sounds rather obvious, however the best definition of a system's usability comes from the end user, and whether they'll drop a current system in preference for a new one. Previous attempts at smart Lab-Books both from commercial [21, 22, 23, 24], and Research Projects [25, 26, 27], have led to systems that are poorly regarded within the community[19]. In many cases this can be attributed to the interfacing of the system. They are generally based on the user's desktop PC, which doesn't translate well into the research laboratory where there is rarely the free space for computer systems, unless they are directly linked to the experiment. In their study Schraefel et Al [19] attributed the success of their system to an understanding of how the chemist works with their lab book through an analogy based design process. They recognised the "hug-ability" of paper lab-book, and embodied this in a wirelessly networked tablet PC. This has the benefits of being portable (is of similar size to an A4 notebook), allows for "freeform" sketching and note taking, and has the remote storage benefits, by being always "online" and connected to the back-end database. By regularly committing data to the data store, the chemist is able to

recover from both software and hardware failure with minimum data loss.

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, this value can be transferred with the sample throughout its life-cycle. When a spectroscopic technique is performed on the sample, by using its unique ID the spectra could be automatically filed into the samples record. When a material is taken on as the precursor to another, by using its unique ID 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 chemists supervisor needs to be consulted about a problem a chain of association could be set up (Chemist is member of Group, Supervisor leads group), hence the supervisor could access the spectroscopic data, along with the experimental observations securely, and easily with indexing being provided in a coherent fashion. 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 you 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 more often 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. It is far too easy to assume that a set of "standard" conditions and apparatus configurations was used. However when revisiting this data later, trying to trace what these "standard" conditions were becomes increasingly difficult. It is almost 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.

3.6 Publish@Source, Publish in Full.

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

Frey and Hursthouse[10] note 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. Taking their example of X-Ray Crystallography, the published data set of structures is currently in the region of 300,000 which is collated in the Cambridge Structures Database[28], however they estimate that approximately 1.5 million structures have been successfully resolved, out of a potential dataset of 10 million compounds. The disparity between the 1.5 million structures known, and the 300,000 published is their major point of concern. They propose 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 it's 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 all the raw data from the analysis techniques carried out on a compound, especially where there exists a possibility for doubt in it's processing as this allows the end user to re-process the data, hopefully, achieving the same results.

4 Experimental

4.1 Data System Overview

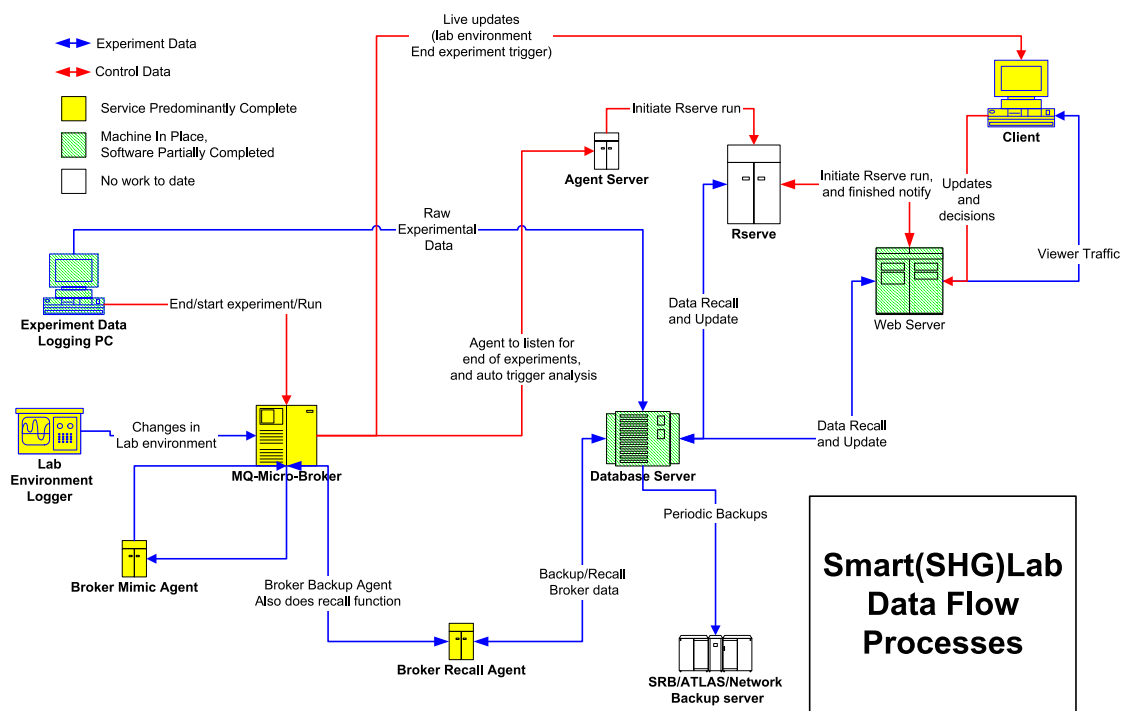


Figure 1: SmartLab Data Flow Overview.

Figure 1 shows all the machines that are expected to be included in the Smart-Lab. In the following sections we break this down to describe the key data paths. There is some system re-use within the SmartLab, this allows for more efficient machine use. The two main examples of this are the web-server, and database servers. These serve both the environmental data flow, and the experimental data flow. The majority of the services are hosted by the same physical machine (MQ server, Agents, Web-server, Database, R-Service), however all communication is done using TCP/IP sockets, hence they could be separated. This would be particularly useful if other laboratories were linked into this system, and the load was found to be too great for the one machine. After concerns about system reliability and security, a stateful packet filtering firewall[29] was

implemented using a dedicated machine, and the Linux netfilter[30] software. This was designed to protect the Microsoft Windows based laboratory data capture machines. Similar local firewalls have also been implemented on all public facing Linux machines in the Smart(SHG)Lab. These problems were highlighted by an infection of the Sasser Worm[31]. The speed with which this worm spread resulted in the experiment data capture machine becoming infected before the appropriate security patches could be installed. Since installing the firewall problem like this have not re-occurred, despite a number of machines in other areas falling victim. Some initial teething problems concerning data access were experienced, but these were quickly solved by adjusting the access permissions granted by the firewall.

4.2 SmartLab Environment Data

Through discussion with operators of the SHG experiment it is noted that certain factors and events lead to lower quality data (i.e. larger standard deviation, poorer fit to model, etc). When working in the lab some of these can be spotted by the operator, and in some case re-starting a run can prevent valuable lab time being wasted. As the lab becomes more automated, it is expected that the operators will be spending less time in the lab (which can also improve results) hence inclusion of a monitoring system is important so that the operator is aware of those occurrences that can spoil data. A demo system was implemented using existing and readily available equipment to test the feasibility of inclusion of such a monitoring system into the Smart (SHG) Laboratory. This initial demonstrator monitors temperature, presence of people in the lab (and their movement in and out) and the state of the lab lights (on/off). A sensor connection to the lab laser interlock was also made, as this provides a quick indication of unauthorised access to the lab (which would result in the interlock tripping).

4.2.1 Electronics

Temperature was measured using three semiconductor temperature sensors supplied by RS-Components (Stock No 317-960). These are supplied mounted in a TO-92 casing and provide a linear temperature to voltage output. This signal is then captured using an existing Data Acquisition card (National Instru-

ments LAB PC+) in a re-cycled PC. These sensors measure temperature in three areas of the labs, away from the rig, near the laser power supply and inside the safety covers alongside the input optics.

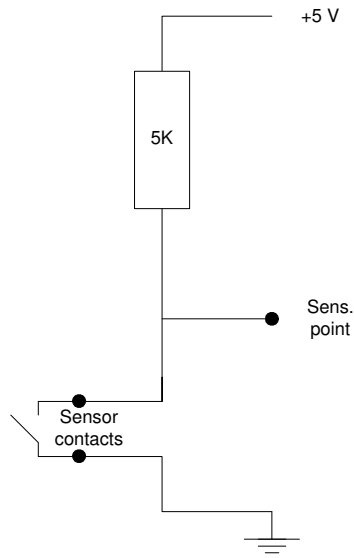


Figure 2: Circuit for the monitoring of toggle sensors.

The state of laboratory light is being 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 toggle that mimics 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 better solution may be to mount a higher sensitivity photo-diode alongside the photo multiplier tube (Optical Output Sensor) on the apparatus. It is expected that the actual level is reasonably constant in the region being studied hence it is adequate to only

record light on/light off. However for an experiment detecting in the visible light region, it may be more appropriate to monitor the actual light level.

Monitoring of personnel movements was performed using security alarm components. There PIR sensors in the lab detect people moving, it has been found that due to the layout of the lab, and the necessity of operators to stay still, that the PIRs lose people in the lab, and stops triggering. This was predicted at the planning stage, so the ability to check people passing through the lab doors was also added. Door state is monitored using simple magnetically controlled reed switches. All these toggle sensors connect back to the TTL Input of the Data Acquisition Card, using the cards internal 5V supply with a bias resistor to generate the high state, as shown in figure 2.

4.2.2 Publish/Subscribe model data broker.

Within the Smart(SHG)Lab project, broker technologies are used to arbitrate a push-based communication system. The broker is used for the publishing of

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

Figure 3: Sample Message for the SmartLab Broker system.

the environment logging data and to inform clients of start/end events on the apparatus. It was decided that the broker will not be used for the transmission of actual experiment data as this is only ever destined for database storage, from which user clients can retrieve it.

Initial work was done sharing an existing MQ Series broker (now WebSphere Business Integration Message Server) from IBM[32] with the FloodNet Project[33]. It was soon found that the extra load and differing messaging content was causing reliability problems for FloodNet, so a dedicated broker was set up for SmartLab. This was based on a developmental Microbroker, written in Java. The microbroker only provides the basic services needed to pass messages, and as such is much easier to configure. In operation it has proved to be very stable. This also has an added data security benefits that the machine is dedicated to this project, hence we can limit its communication to only include those machines needed for the project using firewalling techniques. Data flow within the environment monitoring system is shown in figure 4. The messages being passed follow a loose XML structure, containing an identifier for the sensor, the value being transmitted and a time-stamp (in Unix Time-stamp format). A sample message is shown in figure 3

4.2.3 Storage

It should be noted that data in the broker is only transitional, and that if the data needs to be retained then some external mechanism must be implemented. This was achieved by implementing a simple backup agent, combined with a storage database. The backup agent implements a broker subscribe task, subscribing to the broker tree containing the SmartLab data. On every message reception event the agent inserts an entry into the database, containing the full

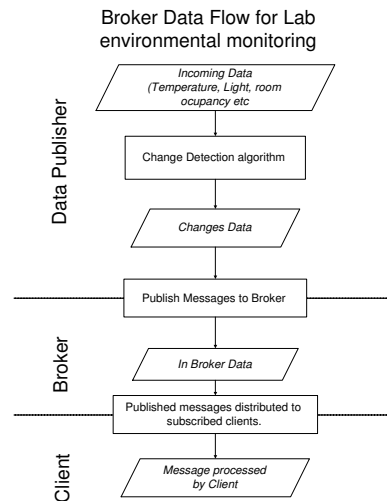


Figure 4: Data Flow from Sensor to Client

broker path, the message time-stamp and the message payload. Once this data is in the database it can be manipulated easily using a number of programming platforms.

4.2.4 Data Recall

Data stored in the database can be easily recalled as it is indexed by time and sensor. Two uses of this are to provide a retrospective inspection mechanism and to provide near real time data display. It should be remembered that there is a slight delay in accessing data in this fashion. To get a true “real time” display a client that subscribes to the broker could be built. A demonstration of “near real time” and “real time” monitoring is hosted at <http://smartlab.combe.chem.soton.ac.uk/smartlab/labdoorframe.html> a sample screen-shot of which is shown in figure 5. This window gives a textual interpretation of the lab sensors in the left side frame, updated once every two seconds. The Java applet in the top right provides a running graph monitoring the temperature sensors, and the bottom right image is a feed from a lab-webcam. At the time the screen shot was taken the webcam showed our office space where the camera server was being configured. It has now been deployed, and two webcam feeds are available, one an overview of the lab, and the second a closeup of part of the output optics.

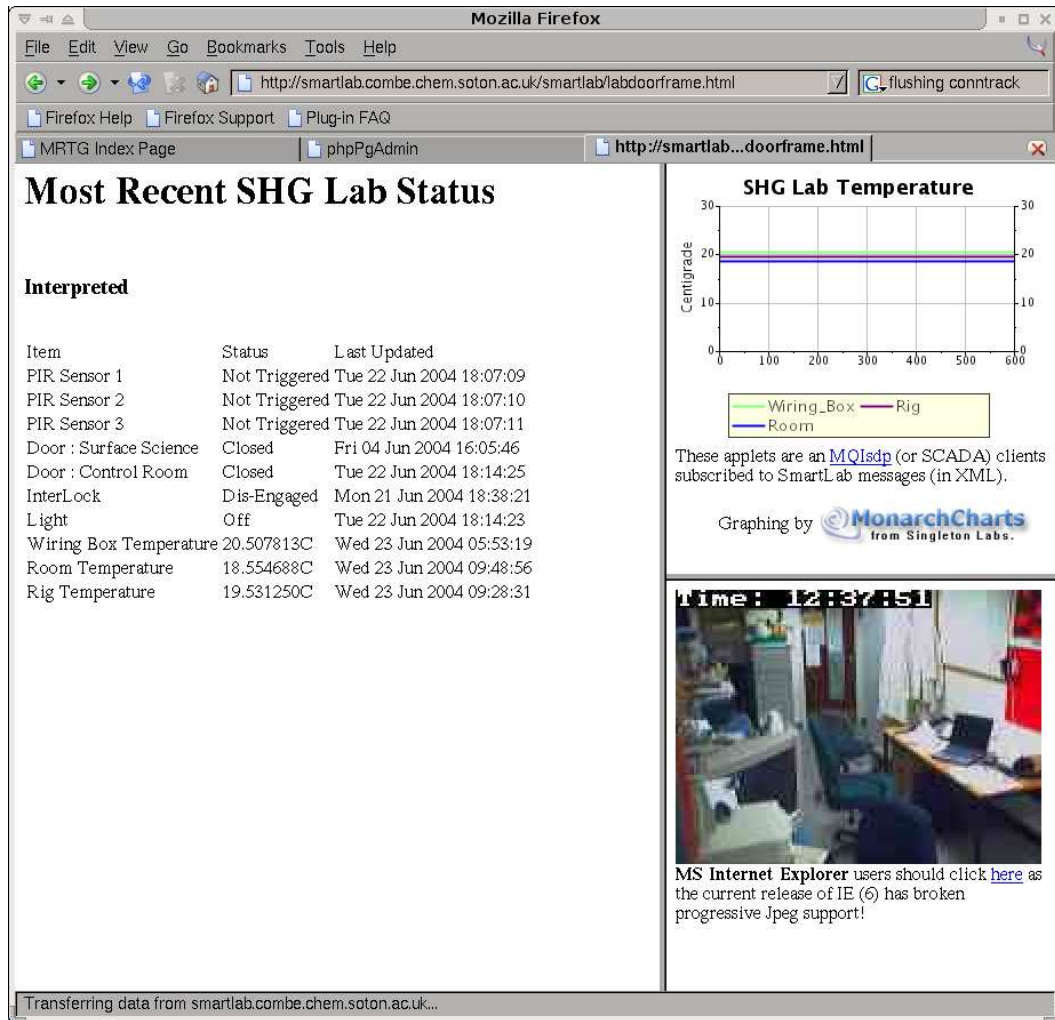


Figure 5: Example Screen-Shot of the lab monitoring webpage.

The Java graphing module available connects directly to the broker as a subscribe client, listening for specified messages. It then plots a point for each message received. The lab sensor capture server only produces messages when there is a change in temperature. In order to get a plot that runs with time a second broker agent was written that subscribes to the lab temperature stream and buffers the data into an internal structure. It then publishes a set of messages (one per sensor) every two seconds on a different stream. The graphing modules then subscribes to this 'mimic' stream, hence plots with time as it is constantly receiving messages.

As mentioned it is also useful to be able to retrieve past data from the environment sensors.

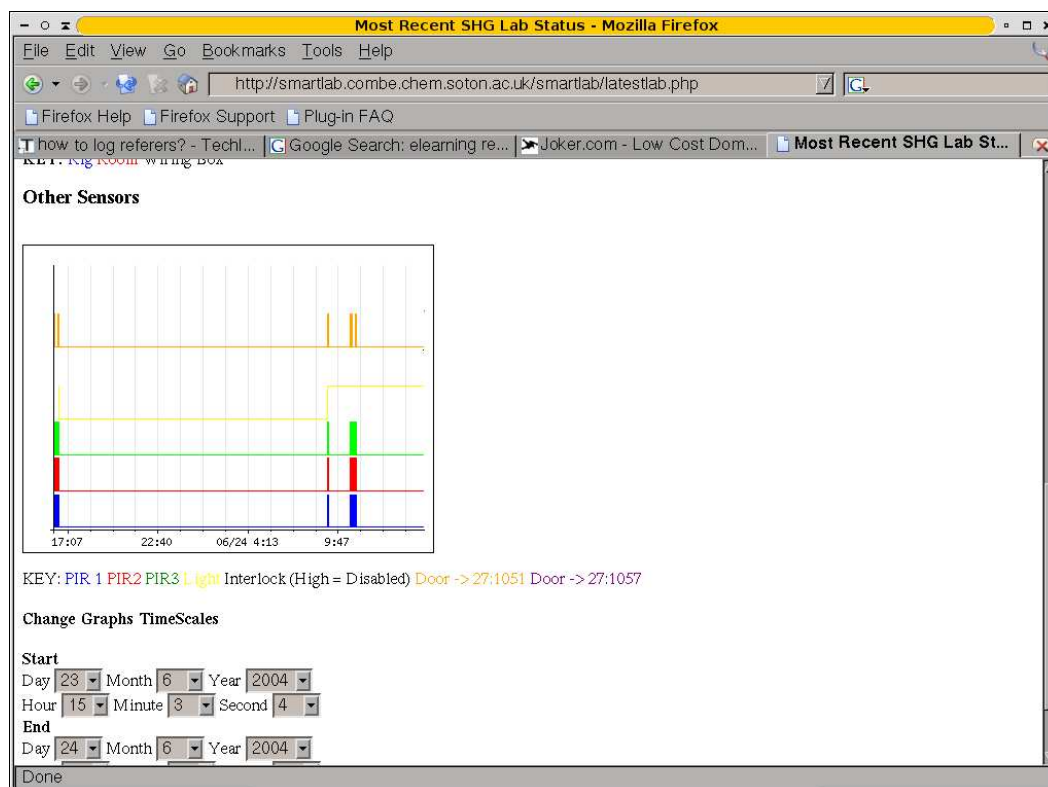


Figure 6: Screen-shot showing part of the "labrecall" page

<http://smartlab.combe.chem.soton.ac.uk/smartlab/labrecall.php> was created as an initial demonstrator of the this principle, the page displays a graphical representation of data for a specified time period (which defaults to the previous 24 hours). The page also gives one the option of re-specifying the displayed time period. Figure 6 shows part of this page, in this case showing the activity of the "toggle" sensors in the lab (eg doors and PIR detectors). This alone is only of curiosity benefit. The next step for this is data is to include experimental data alongside the lab data. In this way it is hope to be able to spot the cause of some anomalous data points.

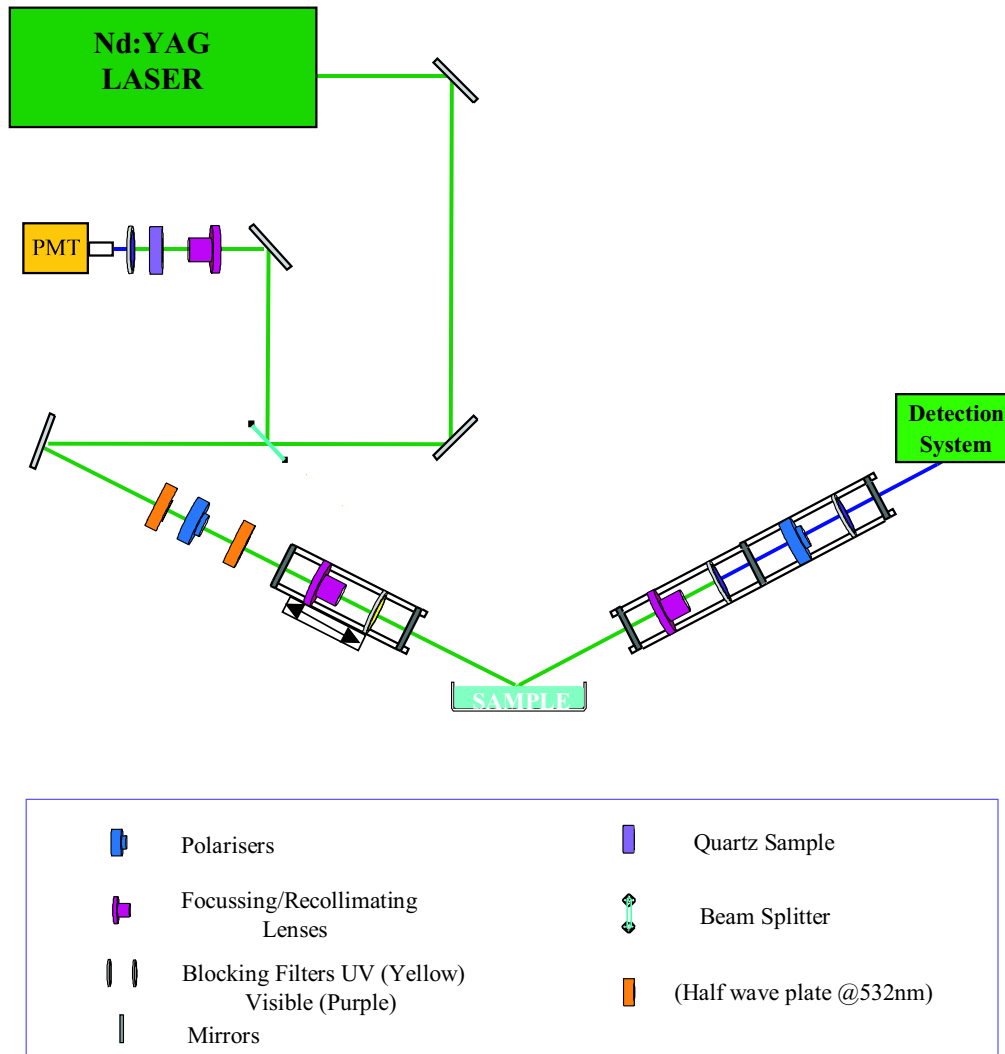


Figure 7: Optics of the SHG Experiment

4.3 SHG Experimental

4.3.1 Optics and Electronics

The SHG apparatus used was as per figure 7. The outputs from the photodiode, and photo-multiplier tube were processed by boxcar gate 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, with a locally built software interface. This controlled the input polarisation, and performed the initial data averaging although the raw data was also available if required.

Monochromator scans were performed by scanning the wavelength of the monochromator, and collecting data samples of 1000 laser pulses at each wavelength. A mean of the data collected at each wavelength was then plotted. Background readings were taken, by shuttering the laser, and collecting a data set of 1000 laser pulses. The mean background value was subtracted from the each data point. Polarisation Scans were performed by adjusting the input polarisation using the half-wave plate in the input arm, under control of the data collection system, then collecting data for 1000 laser pulses and averaging at each input polarisation. Both Monochromator scans, and Polarisation dependence experiments were performed for output polarisations of 0° , 45° and 90° . This was selected using the rotating polariser in the output arm.

4.3.2 Sample Studied

It was chosen to study Rhodamine 6G as this is expected to produce SHG signals at low concentration, but as concentration increases it fluoresces at wavelengths near to the SHG signal. This provides a technical challenge in order to gain good separation of the SHG signal from the fluorescence. It is hoped to be able to demonstrate this by scanning across a range of wavelengths around the SHG wavelength (266nm).

Rhodamine 6G was purchased from Aldrich Chemical Company, and was used without further purification. De-ionised water was used for all chemical operations. Solid samples were weighed using a Mettler Toledo balance to 0.1mg accuracy. Liquid samples were measured using new Class B Volumetric Flasks, and dilutions made using Gilson pipettes, calibrated against the solid sample balance. The spectrometer cell was cleaned using the following procedure, scrub with a general detergent (TeepolTM), Rinse with tap water, rinse with Dilute Nitric Acid, rinse thoroughly with de-ionised water (until conductivity diminishes below the sensitivity of our meter). This procedure is believed to remove all organics, and all significant inorganic contaminants. Sample preparation apparatus was cleaned using the same method.

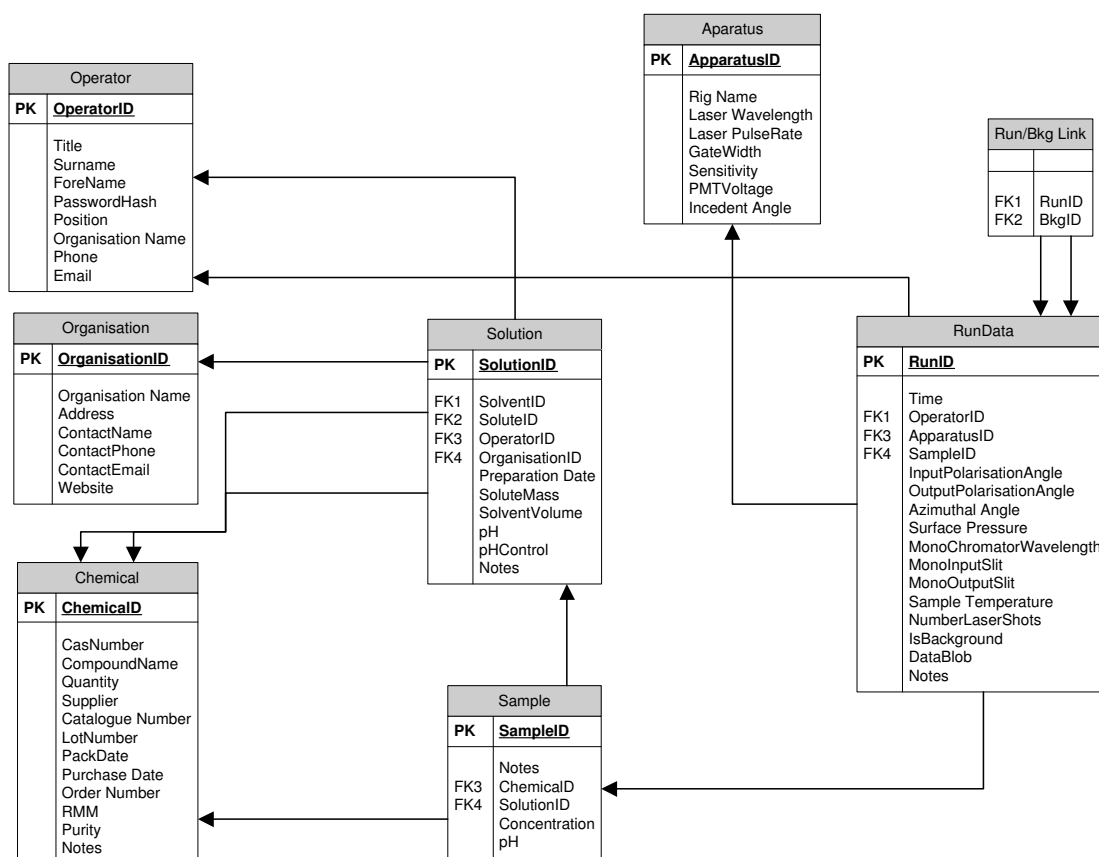


Figure 8: Relational Database Scheme for storage of SHG related data.

4.3.3 Software

To support the Data Acquisition and Processing of data from the Second Harmonic Generation Laser Experiment within an *e*science Framework, it required that a central data-store exists to hold the raw data from the experiment. It was decided to provide this using a relational database, because the parameters involved are reasonably static, and hence generating a rigid schema possible.

Design of the database schema was completed collaboratively with other members of the Frey group who will be involved in it's use, both experimentally and in producing software. 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 extends from the raw data collected by the acquisition card, through the experimental parameters, to information about the sam-

ple being studied, and the chemist performing the experiment. This data was then re-classified by how often it changed (rarely/per sample/per dataset etc) and this used to generate the schema shown in Figure 8.

In order to populate this database, one of two routes can be followed. Either data can be inserted directly into it as it's recorded, or the data can be collected to files, and then retrospectively inserted. In line with pervasive technology ideas, the first of these is preferential, and was the route taken. This also provided the opportunity to re-write the experiment capture software, such that future automation of the rig can be almost trivially included. This was achieved by modularising the design where-ever possible. Some code reuse was possible (interfacing with the data acquisition card for example) but large amounts of code had to be rewritten from scratch.

4.3.4 Managing system inconsistencies

When running experiments, you hope for the condition to be stable for the duration of the experiment. However this is not always the case, and steps must be taken to minimise the effect that this has on the data collected.

Laser Power stability. Laser systems have a finite warm-up period, during which the output power will vary rapidly, until it reaches a stable plateau. For the laser system in use in Southampton this is typically in the region of two to three hours, deviation from this indicates a fault in the laser cooling system. In order to check this an experiment is periodically performed to observe the behaviour of the laser power with time. The experiment involves placing the power-meter sensor in the path of the laser beam, and recording data constantly for the duration of the warm-up period. The sensor is typically placed between the reference beam split, and the first mirror, the analog output being connected directly to the acquisition board, in the place of the PMT boxcar signal. This also allows for the monitoring of peak power over the longer (multiple month) time scale, which indicates the overall "health" of the laser system, and can help to indicate when a service is due. Data is collected using the same software as above, the averaged points then being plotted against time, a best fit line is then drawn in, and an approximately horizontal, linear section identified. This period is the one used for SHG data acquisition, the warm up time is the time between laser startup and the start of this period.

Sample Stability Some samples may be sensitive to air, or be affected photochemically either by the overhead lab lighting, or the laser source. This will cause a change in SHG signal with time, and hence provide a limited data capture window. This window can be identified in a similar way to that used for finding laser stability, although it is typically the period instantly after mounting the sample that is of interest.

The sample would be setup on the experiment in the same way as for polarisation or monochromator scan experiments. A combination of input and output polarisations that gives a good signal would be chosen, typical combinations are 0° input, and 90° output or 45° input and 0° output. Data would then be collected using the usual method, although typically with a much smaller bin size (100 laser pulses per point instead of 1000), as this gives better time resolution, at the expense of standard error. Again a plot of the averages against time is produced, and the linear (near horizontal) periods identified.

Fluorescence, and other two photon phenomena. Some samples (R6G is an example of this) exhibit other phenomena at wavelengths near the SHG signal. These will interfere with the data measured, and may un-necessarily skew the data. It is important to ensure this effect is minimised. This can be achieved by minimising the width of the input and output slits on the monochromator, making it more wavelength dependant. However this is at the expense of signal strength, and as the slits become narrower the alignment of the optics becomes more important. To check the separation of SHG signal from other signals, a scan of the detection wavelength is performed. This is performed at fixed polarisation that gives a strong signal, possible values have already been indicated. Data is typically collected over periods of 1000 laser shots at each wavelength. This should show the SHG signal as a sharp peak with shoulders extending to approximately zero intensity. This test is referred to as a "Monochromator Scan".

Optics zero alignment Computer software or hardware faults can cause the zero position of the automated optics to be lost. The zero point can be confirmed by using a sample with know SHG response. Typically a sample of de-ionised water would be used, as this gives zero SHG response at input polarisation 0° , output polarisation 0° . A series of experiments would be performed, usually

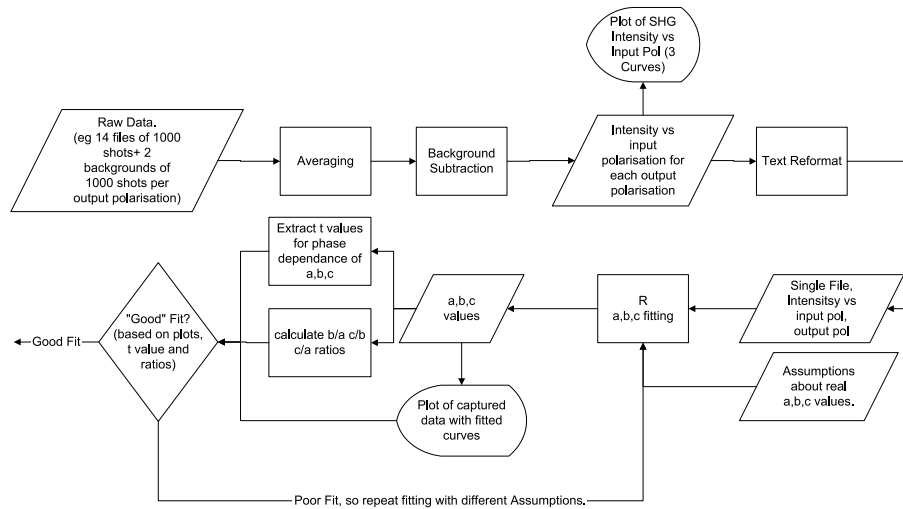
over 10 angles with a lower (typically 500) number of laser shots per angle, starting with a scan over a wide range, then narrowing the scan distance, as the zero point is located.

4.3.5 Analysis (Current) of SHG data.

In order to create the (semi)-automatic SHG analysis and analysis tracking system, one must first have a good understanding of the analysis process that the data undergoes. To this end time has been spent analysing the existing data analysis pathways. The analysis path being followed was put together by Lefteris Danos during his PhD, and as such is documented in his thesis[6]. Following on from discussion with him, and from experience gained so far figure 9 was produced which describe the analysis work-flow for polarisation dependence measurements. It shows the series of analysis processes starting with the raw data from the experiment, through to the generation of conformational parameters.

SHG Polarisation Dependence Analysis.

This experiment measures the Intensity of the SHG signal as a function of input and output polarisation. Throughout the experiment, the sample, power and monochromator settings are kept constant. The experiment may be repeated for other concentrations or powers to produce comparison lines.



SHG Polarisation Dependence Analysis.

This experiment measures the Intensity of the SHG signal as a function of input and output polarisation. Throughout the experiment, the sample, power and monochromator settings are kept constant. The experiment may be repeated for other concentrations or powers to produce comparison lines.

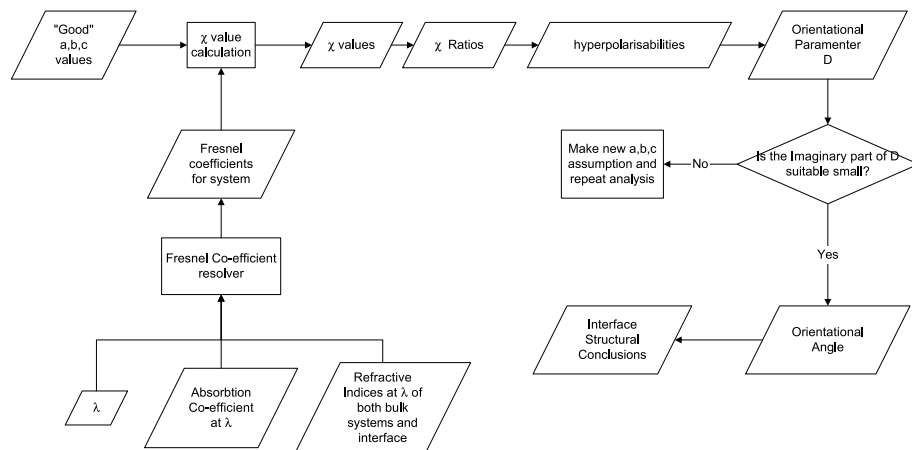


Figure 9: Analysis work-flow for SHG polarisation dependence

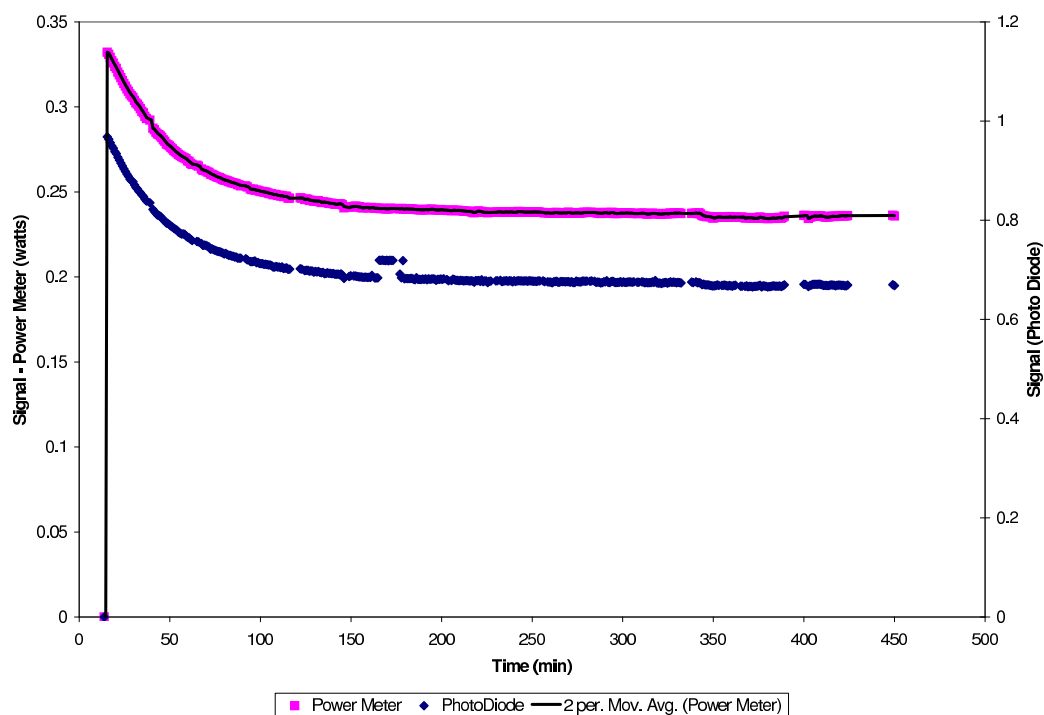


Figure 10: Power Stability of the SHG lab MiniLite Laser.

5 Results

5.1 “System Inconsistency Management” data

5.1.1 Power Stability

Figure 10 was taken at a time when the laser cooling system was suspect, hence the relatively long warmup time of about 3.5 to 4 hours.

5.1.2 Signal Stability

Figure 11 shows a stability scan for a 1×10^{-8} M solution of R6G in water. This scan was performed after being unable to get reproducible polarisation curves. It shows that this particular sample was displaying an increase in signal with time, after a relatively stable 30-40 minutes.

Figure 12 shows the stability for a 5×10^{-6} M solution of R6G in water. This concentration is one of the strongest concentrations that can be comfortably handled in this experiment. The corresponding Monochromator scan is shown in

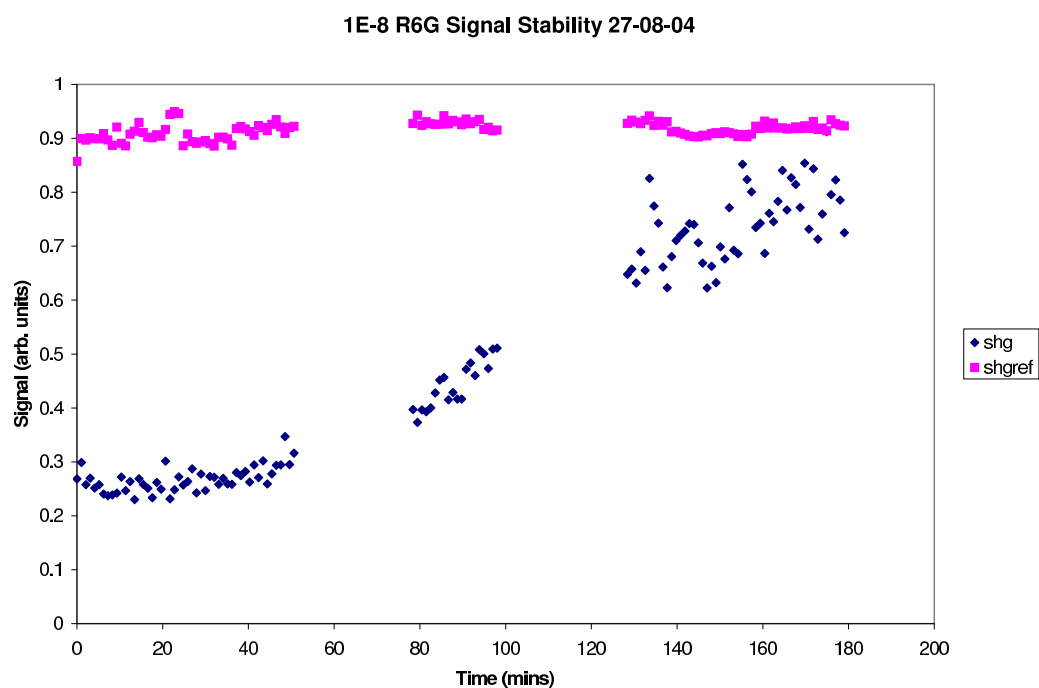


Figure 11: Scan to test the stability of a weak R6G signal.

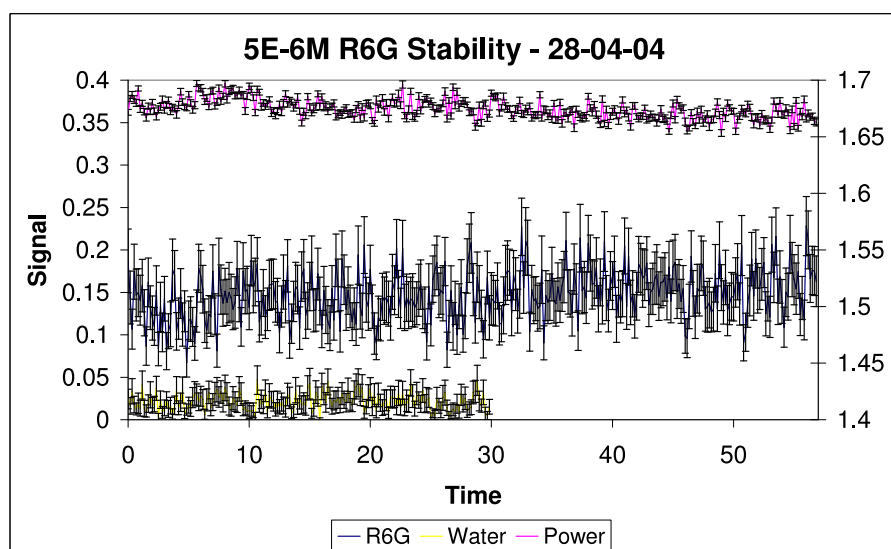


Figure 12: Stability plot for a 5×10^{-6} M R6G solution.

figure 13 which shows the fluorescence problems experienced with this sample. It was left exposed to the air, as there was some prior experience of it oxidising, which would affect the SHG signal. The plot shows however, that within the bounds of the error bars the signal appears to be stable throughout the time period shown (1 hour).

5.1.3 Monochromator Scan

Figure 13 shows a monochromator scan for a 5×10^{-6} M solution of R6G in water. As this graph shows, there is a large observable fluorescence that almost swamps out the R6G peak. This graph shows a non-ideal situation. Figure 14 shows a much better separation, in which the SHG peak is large compared to the fluorescence. This sample is a slightly higher concentration (1×10^{-5} M R6G in water), so demonstrates well the separation that we can achieve under ideal conditions.

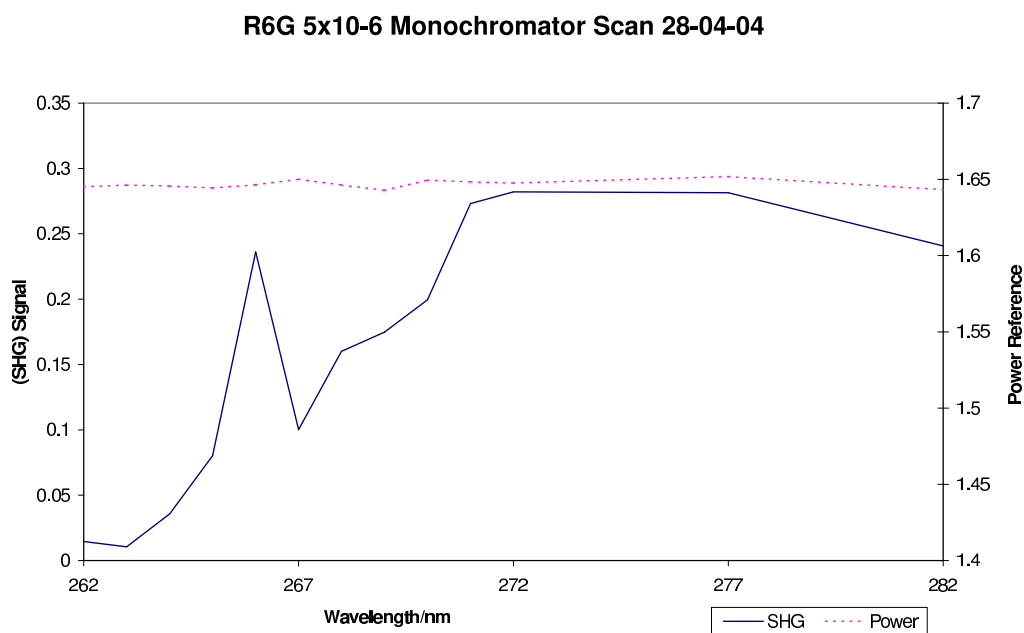


Figure 13: Monochromator Scan showing the strong fluorescence peak observed in this sample.

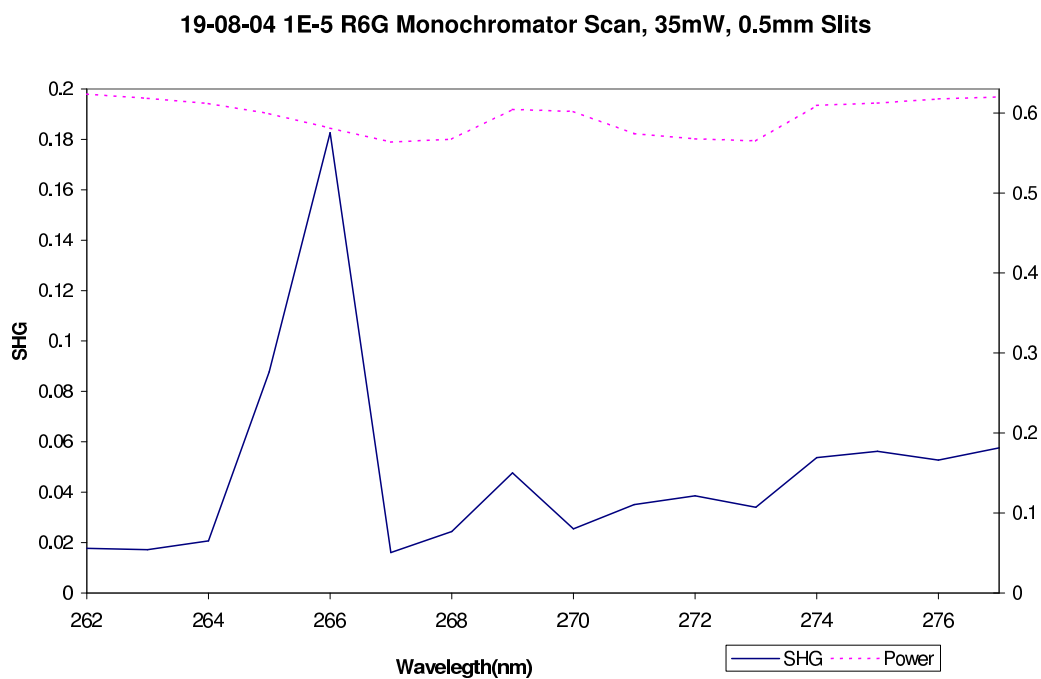


Figure 14: A more ideal monochromator scan, showing good separation of the SHG peak from the background.

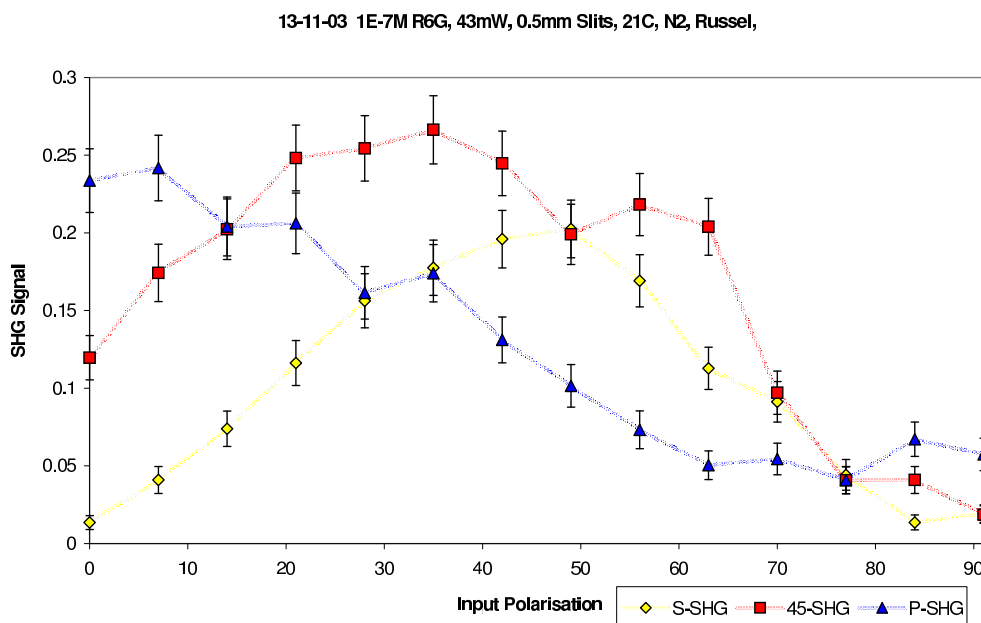
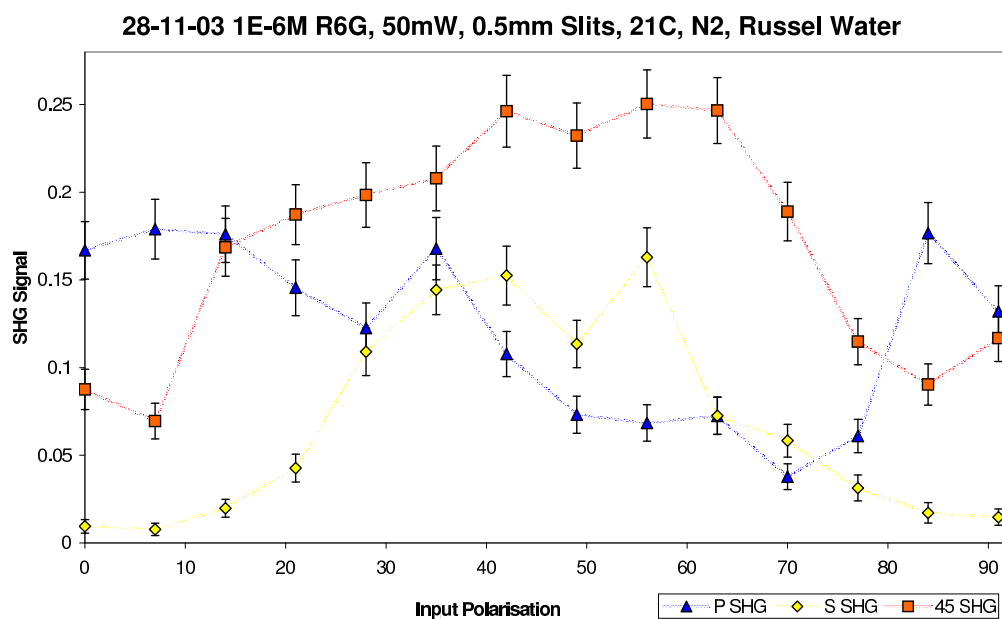
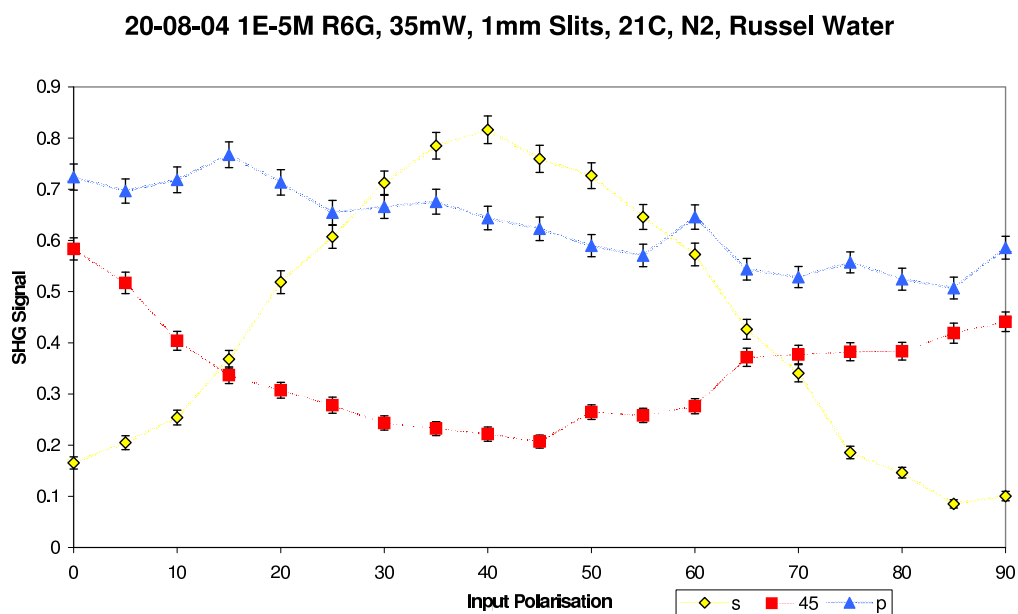


Figure 15: Polarisation dependence scan for 1×10^{-7} M R6G in water

5.2 Raw data

Figures 15, 16 and 17 show a sample of the R6G polarisation data collected to date. Data also exists for 1×10^{-8} M R6G solutions, but it is not included here, as it's hasn't been reproduced yet. The reason for the irreproducibility is shown in section 5.1.2 where I discuss the measurement of signal stability. From the data collected, the general shape of the 1×10^{-8} M R6G data appears to be similar to that of the 1×10^{-7} M R6G data (figure 15). From inspection of the 1×10^{-7} M R6G data (figure 15) and 1×10^{-6} M R6G data (figure 16) it can be seen that the same general shape is retained, with some uncertainty near 0° and 90° input polarisations. However when considering the 1×10^{-5} M R6G data (figure 17) it is quite obvious to see that there is a big change in the 45° output polarisation curve, relative to the lower concentrations. It has been observed in other, currently unpublished, experiments within our group that a transition in configuration occurs around this concentration. Work is currently underway to resolve this potential configuration change.

Figure 16: Polarisation dependence scan for 1×10^{-6} M R6G in waterFigure 17: Polarisation dependence scan for 1×10^{-5} M R6G in water

6 Future Work

- **SHG Case Study**

The smartlab techniques being implemented greatly help to increase the throughput of the experiment. It is hoped to perform a study of a series of SHG active molecules, unlike the single molecules that may have previously formed a paper. By using SmartLab Techniques, all the data will be stored into a central database, ready for publishing. This will lead to an electronic “SmartPaper” to accompany one published in the traditional manner. The SmartPaper will allow hyper linking back to the raw data, allowing the reader to manipulate, and re-analysis of the data.

- **Statistically Driven SHG data collection**

Work already started with member of the Statisticians in the School of Mathematics is leading towards more effective ways of collecting data. By performing a statistical analysis of the modelled response one can highlight the “Interesting” parts of the model, and then concentrate data capture here. Of course some samples won’t comply to the models, hence it is also important to perform initial scans of the whole data collection scope as well.

- **Real time video monitoring**

Initial work involving low resolution “WebCams” has proved to be useful, by allowing the scientist to visually monitor the lab remotely. However the image quality of the current camera is very poor, and hence the useful scope limited. It is planned to collaborate with Researchers in the School of Electronics and Computer Science on the use of high resolution video streaming technologies. These higher resolution images may open the path for some automated image processing, for example to check the alignment to key parts of the rig, throughout the data capture process. This would lead to having a higher confidence in the data presented.

- **Further Design iteration based on data collected in case study, and then performing a further case study.**

Depending on the time taken to perform the above mentioned case study, it may be appropriate to repeat the design cycle once completed, leading

the way for future improvements to the experiments performance. Alternatively it may be appropriate to apply the techniques learned with the Smart-SHG-Lab to other experiments.

- **Further Automation of the experiment to improve efficiency**

At the moment there are a number of areas of the experiment that could be easily automated, and some which would be harder to achieve. The initial aim is to motorise the output polariser, and provide a computer controlled beam stop. This would allow the system to perform a complete polarisation dependence experiment without human intervention. When dealing with compounds such as R6G which have additional emissions near the SHG frequency, it is important to perform the “Monochromator scan” experiments described earlier. It would therefore be a sensible step to motorise the wavelength selection.

- **Production of Electronic Glossary**

As mentioned in the foreword, this work is of interest to researchers across a number of disciplines. The provision of a glossary of terms is an important tool to aid understanding across the potential readership. As we move towards providing electronic “SmartPapers”, then providing an electronic glossary is important. The glossary text, and references, provided in this document provides a starting point for this, which must be extended as the project progresses.

- **Provision of standardised analysis resources.**

Work is already underway elsewhere in the Frey Group to provide the tools to perform, and record, the analysis of polarisation dependence SHG data. These tools must be integrated into the smartlab system, and extended to cover the full range of SHG experiments performed here. It is important to ensure that they provide publish quality output, as well as working screen previews. Tools must be provided to allow the recall of analysis results and processes for linking into the “SmartPaper”.

7 Glossary

The purpose of this section is to provide a quick reference to aid the understanding of some areas of the report, and where appropriate providing external reference for more in depth understanding. It will be extended to form an on-line reference to support future (electronic) publications in the area.

Some of the glossary entries are verbatim copies, modified copies, or extracts from <http://www.wikipedia.org> and as such are referenced back to them 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 here <http://www.gnu.org/licenses/fdl.html> as such they should be considered as a combined work. All other entries are my own work, and are referenced appropriately, in line with University of Southampton guidelines.

7.1 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_\(physics\)](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

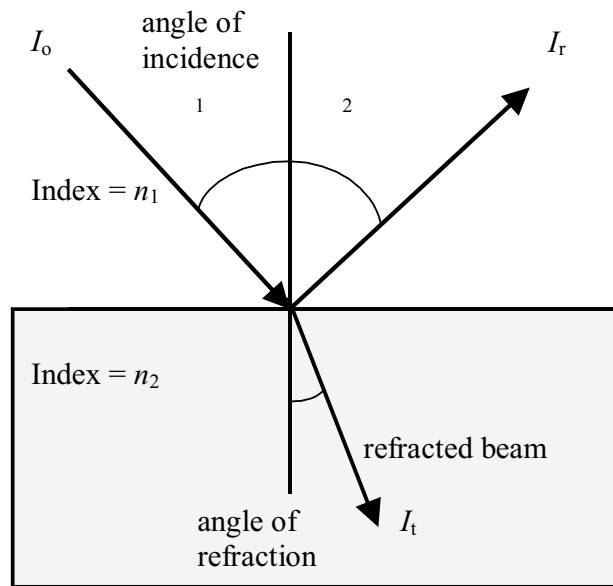


Figure 18: Schematic of light reflecting and refracting from an interface between materials of refractive index n_1 and n_2 . [6]

9. Fresnel Equations

The Fresnel Equations describe the behaviour of light beam when moving between two media of differing refractive indices. At the point the beam meets the interface, reflection and/or refraction may occur. Figure 18 shows this schematically.

The fraction of the incident light that is reflected from the interface is given by the 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 dependant on the polarisation of the incident light. For light with an electric field perpendicular to the plane of figure 18 (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 18, 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 transition 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 nonlinear effects are those which cannot be modelled by a linear relationship.

Nonlinear optics is the branch of optics that describes the behaviour of

light in nonlinear media, that is, media in which the polarisation, \mathbf{P} , responds nonlinearly to the electric field, \mathbf{E} , of the light. This nonlinearity 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 describe 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 their frequency, phase and propagation direction. To fully characterise an electromagnetic wave, the direction of it's electric and magnetic components relative to the direction of propagation needs to be known, it is this relative orientation that defines the lights 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 permeativity 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° , 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

7.2 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[34] 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. It's 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 analog 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, like a home, office or small group of buildings such as 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. MetaData

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 (WRT networked 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 it's networked resource, where as an offline application is one which stores enough data locally to perform it's 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 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. Typically used to replace paper notebooks, as they can provide a similar interface.

17. TCP/IP Network

A computer network using TCP/IP as its logical transport layer. TCP/IP is a hardware independent transport layer, that can be mapped onto a wide range of physical transports. Its most common uses today are as the underlying protocol of the global Internet, and as a transport layer in many 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 and 0.8V, and binary 1 is between 2 and 5 volts. 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 (pronounced as "earl" or spelled 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_\(computing\)](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, eg <Marker>Data</Marker>.

For more details see <http://en.wikipedia.org/wiki/XML>

References

- [1] K. B. Eisenthal, "Liquid interfaces," *Accounts of Chemical Research*, vol. 26, pp. 607–650, December 1993.
- [2] P. W. Atkins, *Physical Chemistry*, ch. 6, p. 154. OUP, 6 ed., 1998.
- [3] K. B. Eisenthal, "Liquid interfaces probed by second-harmonic and sum-frequency spectroscopy," *Chem. Rev.*, vol. 96, no. 7, pp. 1343–1360, 1996.
- [4] R. M. Corn and D. A. Higgins, "Optical second harmonic generation as a probe of surface chemistry," *Cehm. Rev.*, vol. 94, pp. 107–125, 1994.
- [5] P. B. Miranda and Y. R. Shen, "Liquid interfaces: A study by sum-frequency vibrational spectroscopy," *J. Phys. Chem. B*, vol. 103, pp. 3292–3307, April 1999.
- [6] L. Danos, "Non-linear spectroscopic studies at interfaces: Experiment and theory." PhD Thesis University of Southampton, September 2003.
- [7] M. J. Crawford, "Second harmonic generation from liquid interfaces.." PhD Thesis University of Southampton, January 1995.
- [8] V. Mizrahi and J. E. Sipe, "Phenomenological treatment of surface second-harmonic generation," *J. Opt. Soc. Am. B*, vol. 5, pp. 660–667, March 1988.
- [9] J. G. Frey, *Comb-e-Chem: an e-science research project*, pp. 395–398. Blackwell, 2002.
- [10] 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.
- [11] "Clef - clinical e-science framework." <http://www.clinical-esience.org/>.
- [12] "Biosimgrid - a grid database for biomolecular simulations." <http://www.biosimgrid.org/>.
- [13] "Dame - distributed aircraft maintenance environment." <http://www.cs.york.ac.uk/dame>.

- [14] "Coakting - collaborative advanced knowledge technologies in the grid.." <http://www.aktors.org/coakting>.
- [15] "Mygrid: Directly supporting the e-scientist." <http://www.mygrid.org.uk/>.
- [16] "Combechem - structure-property mapping: Combination chemistry and the grid." <http://www.combechem.org/>.
- [17] "Realvnc." <http://www.realvnc.com/>.
- [18] A. Stanford-Clark, "Integrating monitoring and telemetry devices as part of enterprise information resources," tech. rep., IBM, 2002.
- [19] M. Schraefel, G. Hughes, H. Mills, G. Smith, T. Payne, and J. Frey, "Breaking the book: Translating the chemistry lab book into a pervasive computing lab environment," in <http://eprints.ecs.soton.ac.uk/8641/>, 2003.
- [20] A. Dix, "Absolutely crackers, designing an experience." <http://www.hcibook.com/alan/papers/crackers2001/>, November 2001.
- [21] "Cambridgesoft corporation, chemoffice." <http://www.camsoft.com>, 2004.
- [22] "Notebookmaker ltd, notebookmaker." <http://www.notebookmaker.com/>, 2004.
- [23] "Scrip-safe security product, inc, scrip-safe." http://www.scrip-safe.com/laboratory_notebooks.htm, 2004.
- [24] "Intellichem ltd - suppliers of electronic lan notebook solutions.." <http://www.intellichem.com>, 2004.
- [25] L. Arnstein, G. Borriello, S. Consolvo, C. Hung, and J. L. Su, "A smart environment for the cell biology laboratory," *IEEE Pervasive Computing Magazine*, vol. 1, pp. 13–21, July-September 2002.
- [26] 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*, pp. 41–50, ACM Press, 2002.

- [27] J. D. Myers, "Collaborative electronic notebooks as an electronic records: Design issues for the secure electronic laboratory notebook (eln)," in *Proceedings of the International Symposium on Collaborative Technologies and Systems*, 2003.
- [28] CCDC, "The cambridge structural database - the world repository of small molecule crystal structures." <http://www.ccdc.org.uk/products/csd/>, 2004.
- [29] "Stateful firewall from wikipedia." http://en.wikipedia.org/wiki/Stateful_firewall, 2004.
- [30] "Netfilter firewalling, nat and packet mangline for linux 2.4." <http://www.netfilter.org/>, 2004.
- [31] Sophos, "Virus information w32/sasser-a." <http://www.sophos.com/virusinfo/analyses/w32sassera.html>, 2004.
- [32] IBM, "Websphere business integration message broker (formerly websphere mq integrator broker)." <http://www-306.ibm.com/software/integration/wbimessagebroker/>, 2004.
- [33] "Floodnet: Pervasive computing in the environment." <http://envisense.org/floodnet/floodnet.htm>.
- [34] M. H. S. G. schraefel m, Hughes G. and F. J., "Making tea: Iterative design through analogy.," in *Proceedings of Designing Interactive Systems*, 2004.