# Microwave resonant sensors

Jack Naylon

A thesis submitted to Cardiff University
for the degree of Doctor of Philosophy

September 2011

# Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree

Signed…… (candidate) Date……

# Statement 1

This thesis is being submitted in part fulfilment of the requirements for the degree of PhD

Signed…… (candidate) Date……

# Statement 2

This thesis is the result of my own independent work/investigation except where otherwise stated. Other sources are acknowledged by explicit references

Signed…… (candidate) Date……

# Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations

Signed…… (candidate) Date……

# Abstract

Microwave resonant sensors use the spectral characterisation of a resonator to make high sensitivity measurements of material electromagnetic properties at GHz frequencies. They have been applied to a wide range of industrial and scientific measurements, and used to study a diversity of physical phenomena. Recently, a number of challenging dynamic applications have been developed that require very high speed and high performance, such as kinetic inductance detectors and scanning microwave microscopes. Others, such as sensors for miniaturised fluidic systems and non-invasive blood glucose sensors, also require low system cost and small footprint. This thesis investigates new and improved techniques for implementing microwave resonant sensor systems, aiming to enhance their suitability for such demanding tasks. This was achieved through several original contributions: new insights into coupling, dynamics, and statistical properties of sensors; a hardware implementation of a realtime multitone readout system; and the development of efficient signal processing algorithms for the extraction of sensor measurements from resonator response data. The performance of this improved sensor system was verified through a number of novel measurements, achieving a higher sampling rate than the best available technology yet with equivalent accuracy and precision. At the same time, these experiments revealed unforeseen applications in liquid metrology and precision microwave heating of miniature flow systems.

# Acknowledgements

This work would not have been possible without the assistance of many people. I would like to extend my heartfelt thanks to my supervisor Prof. Adrian Porch. His insight and encouragement have been essential over these past four years. I am especially grateful for his early guidance onto this scholarly journey, and his wisdom in allowing me to pursue my own path throughout it.

I would also like to acknowledge the advices and patience of Prof. David Barrow, whose laboratory I frequently disordered during the course of my experiments. Other thanks go out to Prof. Allan Belcher, for his understanding in the final phases of my writing up, and to Mr. Steven Watts, for his tolerance of my monopolisation of his hardware for my first multitone system. Many thanks to Dr. Jonathan Lees, who through his generosity showed me that, in helping resolve the problems of others one may often find the solution to one's own. I therefore owe much to *les étudiants de* L'Université Bordeaux, Jérémy Hyvert *et* Damien Lucas, as well as Oliver Squires, Chris John, Sarah Gooding, and other students I have helped supervise.

I would like to acknowledge the financial support of the Cardiff Partnership Fund, who provided the funding for the prototype hardware, and the School of Engineering for my finance and the delightful opportunity to lecture on several fascinating courses over the years of my doctorate. I am also grateful for the guidance of our Research and Commercial Division, and to Dr. Glesni Pierce and Dr. Paul Goodwin in particular.

Many thanks go out to my friends and colleagues over the years, especially Dr. Oliver Castell, who was an inspiration, and for the mutual help and support of David Rowe, Daniel Slocombe, Alex Morgan, Alan Clarke, *et al.*, and to the jolly good chaps Jonathan Chong and Aled Stephens, as well as many other neglected friends. I owe a special thanks to my partner Zuzana Mrázková: I could not have done it without your love and understanding. Finally, I am eternally grateful to my mother Wenslie and father John; I wish only that he could have been here to see his son complete his PhD, just has he himself did many years ago.

# Contents

# 1 Introduction

Microwave resonant sensors (MRS) exploit the change in the spectral response of a resonance to make enhanced sensitivity measurements of a wide range of physical phenomena linked to the change in a material's electromagnetic properties at GHz frequencies. One such phenomenon is the modulation of the microwave surface impedance of superconductors by impinging light waves. This has been exploited recently to create state of the art detectors for telescopes in the far infrared region of the spectrum [1] – vital for our understanding of galaxy and star formation – as well to engineer better THz cameras [2], recently introduced in some airports to detect concealed weapons.

Whilst the GHz frequency band does not always contain useful information about sample substances, it does not ionise or chemically alter samples, can be applied without direct contact, and penetrates through a large variety of materials. In addition, MRS design is highly adaptable, enabling a host of specialisations from super-ruggedised sensors monitoring oil and gas composition at deep-sea wellheads to nanofabricated microwave microscopes capable of imaging single atoms and probing the localised properties of complex materials such as high temperature superconductors and thin film ferroelectrics. Many other interesting and relevant phenomena can be studied using MRS, and the same basic sensor principle has been tailored to suit a diversity of measurement tasks and conditions.

Since the 1950s, these advantages have attracted a range of established uses for MRS [3], but outside of the laboratory they tend to be limited to less demanding 'industrial' tasks such as density estimation, moisture content measurement and quality control [4]. These applications do not truly harness the power and precision of laboratory-based microwave resonator measurements, which can achieve a very high dynamic range and accuracy but are traditionally slow and require a high level of expertise, laborious calibration routines, and awareness of environmental variability.

Despite this, a number of new and potential uses for MRS systems have arisen that demand fast sample times, high accuracy and laboratory precision as well as presenting complex system integration challenges. These include the aforementioned kinetic inductance detectors for astronomy and THz imaging, as well as scanning microwave microscopes for surface characterisation [5]. Other more speculative applications include precision heating in miniaturised flow systems [6] and sensors for portable instruments and medical devices [7]. The key to enabling or enhancing these applications is changing the architecture of the electronics for spectral characterisation to make it faster, cheaper, and easier to integrate whilst retaining the high performance of laboratory-based microwave resonator measurements. Therefore:

> The aim of this thesis is to solve the issues currently limiting the speed, accuracy, and integration of MRS systems so as to enable their advantages to be brought to bear in new and unforeseen applications.

Fortunately, advances in digital signal processing (DSP) and broadband converter technology created an opportunity to move away from traditional MRS readout architectures. Part of the solution was to use multiple frequencies of excitation simultaneously (a so-called multitone signal) allowing spectral characterisation without requiring local oscillator tuning or invoking resonator transients, giving a big speed advantage. To generate and measure such a signal, I developed a system for resonator spectral characterisation, the readout hardware (Chapter 4).

The basis of the system is a wideband, high precision, software-defined radio (SDR) transceiver, using high-resolution, broadband analogue to digital and digital to analogue converters to cover not only the full spectral response of the resonator but also its intended perturbation range. The design is flexible and has minimal radio frequency (RF) components, instead relying heavily on DSP. In certain frequency ranges aligned to the needs of digital communications, it is fully implementable with just a few highly integrated semiconductor components. This means the system is scalable for low cost, low power consumption and small physical dimensions.

The concept of the multitone readout system was recently found to have been developed in parallel [8]. However, Hermann *et al.* did not appreciate the scope of their invention. Despite stating its capability for high-speed readout, they did not demonstrate any novel or enabling applications, nor did they focus on optimising the accuracy and precision of their system for meteorological quality measurement, as done here.

To investigate the performance of the hardware, I conducted long-term stability measurements. These show that the precision of the new measurement system on short time scales is comparable to that obtainable with a conventional vector network analyser (VNA). Over longer time scales, thermal fluctuations in the resonant sensor itself limit the precision of both techniques. Furthermore, if the sensor can be periodically referenced to a measurement of a known standard (such as air), the achievable precision in realtime resonant frequency measurements is about 1 part per billion, or 30 bits, over a 1 s integration time. This is a higher dynamic range than many other sensor systems can boast.

A large part of this new architecture lies in the algorithms used to calculate the resonator spectral response and extract relevant sensor measurements from it. These algorithms together make up the readout software (Chapter 5). A likely target platform is a field-programmable gate array (FPGA), a type of highly parallel programmable digital integrated circuit, in anticipation of an embedded sensor solution. Therefore, algorithms suited to the highest possible readout speed were developed with parallel processing and high computational efficiency in mind. Extensive use was made of the efficient computation and parallel processing capability of the National Instruments LabVIEW programming environment in order to develop and test these software routines.

The hardware and software elements of this novel MRS readout system are the most significant contributions to knowledge arising out of this work. They are the basis of a patent filed at the end of this project [9]. The desire to protect this intellectual property, necessary to secure funding for the SDR platform used in its

development, has meant that the main body of this research work is yet to be published.

A key part of the attempt to solve some of the other problems facing current MRS applications, such as high systematic error and sensor miniaturisation, was the development of a general resonator modelling method (Chapter 3). Investigation into theoretical sources of nonlinearity in the resonant sensor response identified a need for a comprehensive model, general enough to correct the nonlinear response of complex miniaturised sensors and compensate for the effects of high-frequency dynamic perturbation processes. A combination of theoretical analysis and interpolated finite element method (FEM) eigenvalue simulations provided a hybrid, or semi-parametric multi-stage model. This linked measured resonator spectral response, resonator eigenvalue, and a desired measurand, such a permittivity. Artificial neural networks were proposed as a means to represent these models in a form suitable for FPGA implementation, allowing the sensor electronics to correct automatically for any nonlinear response of the sensor in realtime and at high rate (Chapter 6). Although general and somewhat theoretical at this stage, this work will help support on-going microwave resonant sensor design activities at Cardiff University and could be appropriate for publication if verified experimentally in the future [10].

The development of this readout system greatly enhanced the speed of resonator characterisation over traditional laboratory MRS measurements. This enabled the development of a novel system for controlling solid-state microwave heating in a miniature flow system [6]. I also used the new readout system to measure free-falling droplets and gas-segmented flow streams, primarily as an attempt to verify the accuracy of dynamic MRS measurements using the new readout system (Chapter 7). As well as being largely successful toward this aim, these experiments potentially present a new opportunity to apply MRSs in multi-parameter fluid characterisation – possibly offering a cost or robustness advantage over traditional methods in measuring properties such as surface tension and viscosity, in addition to permittivity [11].

In summary, the original contributions to knowledge are the application of a realtime multitone measurement system to liquid property measurements [11] and to microwave heating control [6] [10], as well as an accurate, efficient, fast, and highly flexible MRS readout system platform [9], suitable for use in the next generation of integrated MRS applications, *e.g.,* [7].

The thesis is broken down into six further chapters: Applications, Theory, Hardware, Software, Verification, and Conclusions. The Applications chapter reviews the industrial and scientific uses for MRSs and gives an overview of the developments and technology in several key areas that lead up to this work. The Theory chapter introduces the principles behind resonant sensors and develops the multi-stage MRS modelling approach. The Hardware chapter discusses MRS readout architectures and the development of the new multitone readout system, along with experiments to assess its static performance in comparison to a VNA. The Software chapter covers spectral estimation using multitone excitation signals, their optimisation, and the correction of distortion effects. This chapter also assesses methods for extracting resonant frequency and bandwidth from spectral estimates. The Verification chapter brings together elements of Chapters 3-5 in several demonstrations of the power and scope of realtime MRS measurements including measurement of liquid droplets, gas-segmented flow streams, microreactor heating control and dynamic capillary filling characterisation. The final Conclusions chapter summarises the achievements of this work and discusses the interesting and groundbreaking potential of its main findings and contributions.

The next part of this chapter introduces microwave resonant sensors in more detail and provides a general discussion of their advantages and disadvantages in comparison to other sensing technology.

## 1.1 Microwave Resonant Sensors (MRS)

Defining what is meant by the phrase ‘microwave resonant sensor’ will help explain the context of the problem. This section discusses the specific characteristics and requirements of the intended application scope encompassed by this label.

A sensor is a device that measures a physical property and converts it into a signal that can be read by an observer or by an instrument. This definition introduces the concept of signals. This also implies that a sensor is a subcomponent of an instrument. The job of an instrument is then to process these signals in a useful way. Therefore, an important part of sensor system design must be deciding what signal processing to do and how to do it in an optimum way. Fundamentally, this is the central topic of this thesis.

In engineering, the term ‘sensor’ has come to imply something that is a subcomponent of a larger system, such as an instrument, or to suggest a level of autonomy or a high level integration of a device into a wider system or network. I use the term in this sense also. The device is perhaps just one small part of a ‘higher measurement objective’, meaning that it and any surrounding system is capable of some degree of autonomy and integration. This type of sensor should not rely on laborious human interaction or calibration in order to produce its output signal, be that for dissemination by man or machine.

In the field of microwave measurements, this distances the intended application space from ‘traditional’ techniques aimed at very high accuracy one-off measurements of pure physical properties. This could be said of the measurement of ceramic dielectric properties or the surface impedance of superconductors, for example. These are not sensor applications, but experiments that happen to use microwave resonators. If the user has to spend an hour carefully setting up the experiment and post-processing the data to make one measurement then he or she can accept this as all part of the experimental method and due scientific rigor. This represents a level of human intervention and time expenditure that is

unacceptable for the applications of MRS discussed here. However, the demarcation is far from sharp. After all a very good sensor could be used in a laboratory setting as well, saving the user time and hassle associated with these traditional techniques. This definition of a sensor goes some way to justifying the pursuit of this problem, and represents some of the objectives of this work.

The term microwave refers to the portion of the electromagnetic spectrum having a wavelength between about 1 meter and 1 centimetre (of corresponding frequency in free space between 300 MHz and 30 GHz) [12], although the MRS principle of a perturbed resonator is still applicable even at optical wavelengths. The physical quantities that can be sensed relate to the interaction of these waves with matter; some property of matter will affect the propagation of these waves and hence be detected (sensed) by microwaves. Sensors are transducers, *i.e.,* they convert energy from one form into another. In MRS, this conversion is between the electromagnetic energy of microwaves and the kinetic energy of molecules, ions, and electrons. It is often stated that microwave sensors are used to measure electric permittivity, $\varepsilon$ conductivity, $\sigma$ and occasionally magnetic permeability, $\mu$ [13]:

$$\varepsilon = \frac{\mathbf{D}}{\mathbf{E}}, \qquad \sigma = \frac{\mathbf{J}}{\mathbf{E}}, \qquad \mu = \frac{\mathbf{B}}{\mathbf{H}}. \tag{1.1}$$

These are transfer functions (linear input-output ratios) that state how easily a material transmits a displacement current **D**, an electric current **J** and a magnetic flux in response to an electric field **E** or magnetic field **H**. Of these 'flows', **D**, **J** and **B**, only the electric current actually represents the physical transport of matter. Each of these parameters is a function of the frequency at which the field alternates – the material's response spectrum – and sometimes the direction of the field as well, known as anisotropy. Anisotropy is present in most regular crystalline materials. A microwave resonant sensor can be designed to measure unambiguously any one of these three parameters at a time, and by suitable orientation of sample with the modal field distribution, also extract anisotropic responses as well [14].

The microwave-matter interaction is a subject of great depth and much interest but it does not form a central part of this thesis because the main aim is very general. Although specific examples or case studies may require knowledge of the microwave-matter interactions in that application, it would be impractical to discuss the nature of all the interactions of all the possible sensors to which the main development of this thesis could be applied. Furthermore, in many cases, one is not interested in these quantities *per se* but what they represent for the material under investigation – its shape, size, composition, temperature, presence or absence. For example, one of the main applications of 'microwave sensors' (albeit non-resonant) is in the measurement of distance – such as that from a car's rear end to its owner's garage wall. The actual permittivity of the wall itself does not matter in the slightest, as long as it is enough to ensure adequate reflection.

The third yet key term in the central theme of this thesis is 'resonant'. A resonant sensor is one designed to have a clearly defined and sharp oscillation at a specific frequency (a unimodal resonance) or a set of discrete frequencies (a multimodal resonance). When a material in a resonator interacts with its electromagnetic field, the resonator changes its oscillation frequency and its bandwidth – *i.e.,* the range of frequencies over which it will resonate best. From these parameters, one can deduce the contribution the material has made to the energy storage and energy loss mechanisms within the resonator. The reason to add this extra complication, rather than simply putting a material in an electromagnetic field and measuring the change in field amplitude, is twofold. Firstly, the resonator amplifies changes in its internal fields due to the sample at frequencies near its resonant modes, thus increasing its sensitivity. Secondly, the resulting change in amplitude and phase of signals close to resonance is coherent, whereas any noise is (ideally) not. Thus, in signal processing terms, the resonator provides a signal to noise ratio (SNR) gain by producing a correlated 'response' across a range of frequencies – usually several times the resonator bandwidth. The trade-off for this gain is that each resonant mode can only be used to measure the response of a material at one frequency. However, this limitation is not as drastic as it may seem. For fundamental physical reasons, spectral features in the dielectric response are normally extremely wide – spanning a decade or more in frequency. This inherent

'smoothness' present in the majority of material responses in the microwave band means that response spectra can be 'pieced together' from multimodal resonators, or indeed multiple resonators, without sacrificing much in the way of detail.

## 1.2 Advantages of the microwave interaction

Apart from in situations when one is specifically interested in a material property at microwave frequencies (to assess its performance as a microwave circuit substrate material, for example), the advantages of using the microwave part of the electromagnetic spectrum are relatively subtle. It is fair to say that these sensors enjoy less widespread applicability than those at longer wavelengths (low frequency electrical measurements) and at specific shorter wavelength bands (optical and near-optical measurements). This is for both practical and physical reasons. From a practical point of view, measuring and generating voltages and currents at lower frequencies is easier and cheaper, because wave phenomena are insignificant on the scale of physical devices and the properties of doped silicon at low frequencies are more favourable for making integrated devices.

At optical wavelengths, cheap, miniature and efficient solid state sources such as light emitting diodes (LEDs) and semiconductor lasers are widely available, as are detectors (photodiodes, CCDs, etc.). Furthermore, the wavelength is small enough that the ray approximation is valid on the scale of most physical devices, making design simpler; antennas are not required for free and guided propagation of optical signals and the optics are highly miniaturised. Specific bands can also benefit from the low cost of components, especially those used for digital communications.

From a physical point of view, electronic and ionic conductivity are linked to a wide range of physical processes (Figure 1.1), yet tends to decrease in significance in the microwave band as it is masked by the displacement current. The ‘skin depth’ in which currents may flow in conductors is also larger at lower frequencies, meaning that bulk conductance properties can be measured. At optical wavelengths, with photonic energy about 1 eV, electronic transitions are the predominant source of electromagnetic (EM) interaction; these are accurately predicted and readily interpreted by quantum theory. The fact that most materials have electronic absorption bands at or near optical wavelengths implies that this band contains a rich source of detailed information about chemical composition,

something that cannot be said about the microwave band. This is one of the reasons why evolution has led the animal kingdom to adopt visible light as its wavelength of choice for its most powerful sensor – the eye.

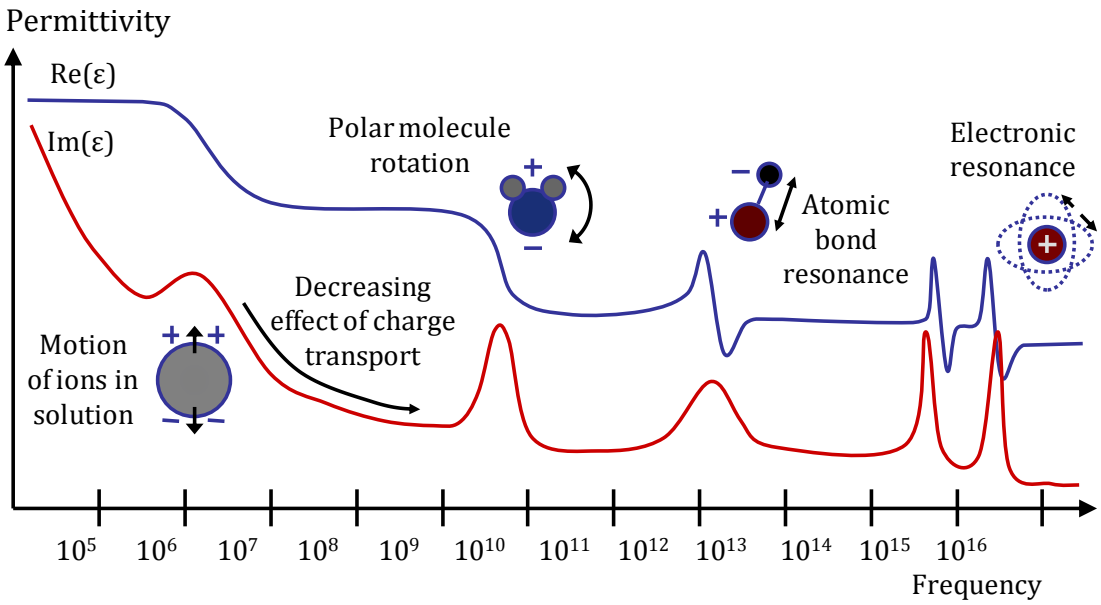



*Figure 1.1: Hypothetical liquid response to electromagnetic fields as a function of frequency, showing the physical significance of various regions of the spectrum.*

Due to its short wavelength, the near-optical band is also much more sensitive to surface disturbances on the molecular scale, such as selective adsorption of analyte molecules into evanescent field regions. This can be further enhanced by the use of surface plasmon resonance; the excitement by laser of surface propagating waves (with associated quantum quasi-particle called the polariton) in a thin layer of metal atoms. This is a primary technology of the latest generation of miniaturised biosensors [15], [16]. The microwave interaction, with photons of less than 1 meV, does not have a quantum interaction (except in unusual cases, such as magnetic resonance) and instead tends to be a combination of many factors, such as density, viscosity, temperature, etc., as it is largely determined by intermolecular structure. This means the response is less readily interpretable and suffers from a multiplicity of perturbing environmental factors. The much larger wavelength also means less selectivity to molecular scale surface effects, as even evanescent microwaves tend to have an active extent many orders of magnitude larger than

that achievable at optical wavelengths. Whilst extreme engineering can improve this resolution to nm length scales using micro/nanofabricated ultra-sharp 'microwave microscope' probe tips [17], this is difficult and costly to achieve.

Despite this, microwave sensors do have many advantages. One often cited advantage is that of being non-contact. Whilst this is true of optical sensors as well, microwave wavelengths are ideally suited to being non-contact whilst also permitting volumetric measurement to a reasonable depth in a wide range of non-metallic materials, whereas non-contact optical methods can only make volumetric measurements in a more limited class of high transmissivity materials. Lower frequency measurements require electrical contacts to be very close or in direct contact with material, so they are less suitable. Microwave radiation is also minimally invasive; it is non-ionising and has high safe exposure limits to operators. Furthermore, in the vast majority of cases microwave radiation cannot affect chemistry except by heating (although this issue has been subject to some controversy in the chemistry community) [18].

A significant advantage is the highly flexible design and construction of microwave resonators; there are many known geometries of resonant structure and although each generally has a disparate 'optimum' geometry in terms of sensitivity, accuracy, size, and ease of manufacture, a good of compromise can usually be reached without sacrificing overall performance. In addition, (unlike optical systems) this flexibility, as well as the relaxed mechanical tolerance when dealing with longer wavelengths, means that the fabrication of microwave sensors does not necessitate expensive specialised or extreme precision manufacturing methods. These advantages imply the ease of MRS integration into extreme environments, complex systems, and cost-sensitive applications.

Cost is a key advantage of using microwave sensors, or more specifically precision-cost ratio. As the sensors themselves are highly flexible in design and do not necessitate specialised materials or fabrication methods, they can be mass-produced at low cost. Thus the significant factor in overall expense is the readout system. The enabling factor for high precision-cost ratio is therefore the fact that

microwaves can be accurately detected and measured by synchronous receivers. Microwave receivers are based on the principle of downconversion by frequency mixing, a process that produces coherent lower-frequency signals that can be sampled with great precision and provide full vector information of a microwave signal. Due to the performance demand of radar and the mass demand of communications over more than a century of engineering, this detector of choice for microwave radiation can be designed with a higher dynamic range than any other room-temperature detector and can be made cheaply enough to be integrated into disposable 'RFID' tags costing a fraction of a penny. Maximising this key advantage, the *precision to cost ratio*, is therefore crucial in allowing the potential of microwave sensors to be fully exploited.

This chapter has outlined the reasons for pursuing the improvements in MRS sensor processing and some of the achievements made over the course of this work. In the next chapter, the applications of MRS will be discussed in more detail, from the more mundane industrial applications that rely on the robustness and large penetration depth of microwave sensors, through to a discussion of some of the interesting new applications alluded to in this introduction, such as kinetic inductance detectors and sensors for miniature fluidic systems.

## 1.3 Chapter references


[1] S. Doyle, P. Mauskopf, J. Naylon, A. Porch, and C. Duncombe, "Lumped Element Kinetic Inductance Detectors," *Journal of Low Temperature Physics*, vol. 151, pp. 530-536, Jan. 2008.

[2] J. Schlaerth et al., "A Millimeter and Submillimeter Kinetic Inductance Detector Camera," *Journal of Low Temperature Physics*, vol. 151, pp. 684-689, Jan. 2008.

[3] A. R. Von Hippel, *Dielectric materials and applications*, vol. 2. Artech House on Demand, 1995.

[4] E. Nyfors and P. Vainikainen, *Industrial Microwave Sensors*. Artech House, 1989.

[5] T. Wei, X.-D. Xiang, W. G. Wallace-Freedman, and P. G. Schultz, "Scanning tip microwave near-field microscope," *Applied Physics Letters*, vol. 68, no. 24, p. 3506, 1996.

[6] J. Naylon et al., "Efficient microwave heating and dielectric characterisation of microfluidic systems," in *Proceedings of MicroTAS*, 2010.

[7] A. Porch and J. Beutler, "Non-invasive blood glucose meters," Wellcome Trust transfer award, 2011.

[8] T. Hermann, G. R. Olbrich, and P. Russer, "A Novel System for Real-time Measurement of the Electrical Properties of a Cavity Resonator," in *Microwave Conference, 2008. EuMC 2008. 38th European*, 2008, pp. 67-70.

[9] J. Naylon and A. Porch, "UK patent application number GB1109685.6: Apparatus and method for estimating a characteristic of a microwave resonant device," 2011.

[10] J. Naylon, "Realtime microwave resonant sensor readout system for flow chemistry applications [to be published]," *Sensors and Actuators A: Physical*, Expected 2012.

[11] J. Naylon, "Measurement of liquid droplet oscillations with a realtime microwave sensing method [to be published]," *Journal of Applied Physics*, Expected 2012.

[12] D. M. Pozar, *Microwave engineering*. J. Wiley, 2005.

[13] C. Kittel, *Introduction to solid state physics*. Wiley, 2005.

[14] G. Annino, M. Cassettari, M. Fittipaldi, and M. Martinelli, "High frequency single-mode resonators for EPR spectroscopy enabling rotations of the sample about two orthogonal axes," *Journal of Magnetic Resonance (San Diego, Calif.: 1997)*, vol. 176, no. 1, pp. 37-46, Sep. 2005.

[15] A. Babajanyan, J. Kim, K. Lee, R. Khachatryan, and K. Nerkararyan, "Glucose concentration monitoring using a surface plasmon polariton," 2008, pp. 672-677.

[16] A. Abbas, M. J. Linman, and Q. Cheng, "New trends in instrumental design for surface plasmon resonance-based biosensors," *Biosensors and Bioelectronics*, vol. 26, no. 5, pp. 1815-1824, 2011.

[17] L. Zhang, Y. Ju, A. Hosoi, and A. Fujimoto, "Microwave atomic force microscopy imaging for nanometer-scale electrical property characterization," *Review of Scientific Instruments*, vol. 81, p. 123708, 2010.

[18] A. de la Hoz, Á. Díaz-Ortiz, and A. Moreno, "Microwaves in organic synthesis. Thermal and non-thermal microwave effects," *Chem. Soc. Rev.*, vol. 34, no. 2, pp. 164-178, Jan. 2005.

# 2 Applications

This chapter looks at some of the applications microwave resonant sensors (MRSs) are routinely used for and what advantages they have, in both a broad 'industrial' context, in Section 2.1, and a scientific one, in Section 2.2.

Section 2.3 then goes on to introduce some of the new and challenging things that MRSs have been applied to in the last decade or so, focussing particularly on those that rely on time-domain sensing; so-called dynamic measurement. These applications have been particularly influential in shaping the development of MRS technology, especially in this work. The first is the microwave kinetic inductance detector (MKID), an extremely sensitive detector originally developed for astronomical telescopes. The concept behind this thesis arose from research into the readout systems for MKIDs, so Section 2.3.1 gives a background to these devices and the previous developments in readout technology that inspired this work.

Cardiff University and other institutions have recently directed some attention toward the integration of MRSs with microfluidic and 'lab-on-a-chip' technology. This interesting and promising research area, currently in a difficult adolescent stage, may find low-cost and high-speed MRS technology valuable in a few specific practical applications – so this field and some of the ways in which MRS could be useful within it are introduced in Section 2.3.3 as well.

The final discussion topic, and pride of the microwave resonant sensor community, is the microwave microscope in Section 2.3.4 – an ultra-high-resolution sensor for probing the high frequency properties of matter down to the atomic level. Because of the need to scan the probe mechanically, and the very small perturbations produced, this sensor is demanding of both good dynamic performance and the highest precision any readout technology has to offer.

## 2.1 Industrial applications of MRS

By a huge margin, the most established application of MRSs is moisture estimation. The accurate determination of moisture content (or more precisely the measurement of liquid water relative mass fraction in solids or liquids, known as aquametry) is crucial to a large variety of different manufacturing processes, from construction materials to pharmaceuticals.

Moisture content affects the quality of products in a range of ways. Many products are prone to bacterial, fungal and pest contamination and high moisture content generally exacerbates this. Improper processing resulting in undesirably high moisture content can impair the quality, efficacy and storage life of foodstuffs, drugs and chemicals, and even pose a significant risk through food poisoning, spore inhalation, or structural failure of building materials. Incorrect water content of mixtures before processing also affects the quality of the resulting products; this is especially true of ceramics, cement, and concrete. Simple economics can also be significant – water is relatively dense, and drying materials costs a lot of money (because of water's high heat capacity); this has implications for many bulk materials bought and sold by weight. Thus, there is a big incentive for manufacturers across the industrial spectrum to measure the moisture content of products throughout the manufacturing process. Fortunately, this is an ideal task for microwaves, and is by far the most widely commercialised application of both resonant and non-resonant microwave sensor systems.

The field of microwave aquametry was reviewed recently by Udo Kaaze, author of over 200 publications in the area of material science and the dielectric properties of water in particular [1]. Other well-known reviewers of the field include Ebbe Nyfors [2], who discusses industrial microwave sensors in detail in his book of the same name [3]. Seichil Okamura, gives a historical overview from 1948-1999 [4]. The late pioneer of microwave aquametry Andrzej Kraszewski who cites hundreds of published applications as well as making significant contributions to the field [5], [6] (Figure 2.1). More recently, Klaus Kupfer's book [7] provides a thorough background to the scientific principles and measurement techniques in the wider

field of electromagnetic aquametry, and has a chapter devoted specifically to microwave resonant sensors.

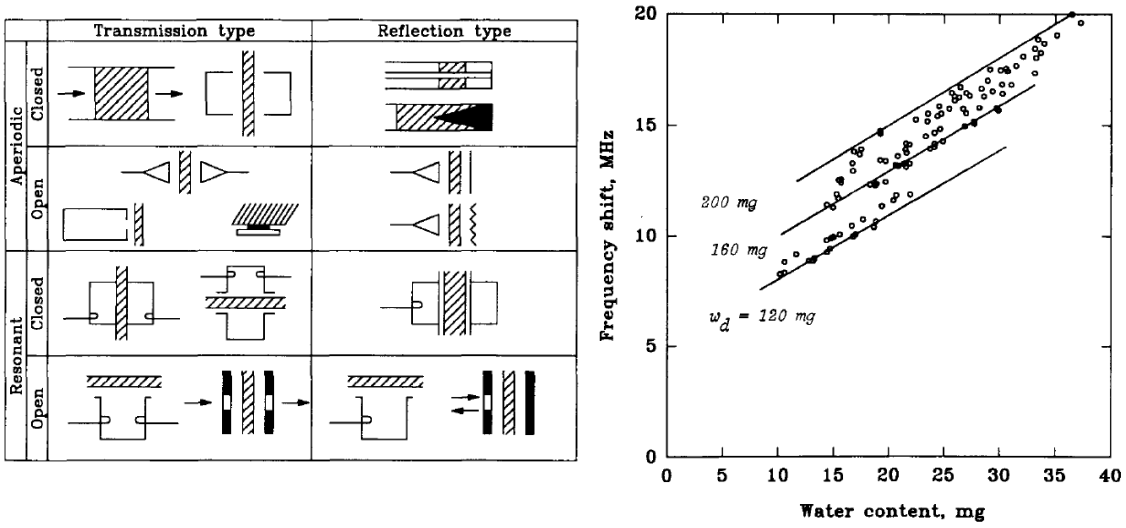



*Figure 2.1: Taxonomy of microwave sensors (both resonant and nonresonant) for aquametry (left) and MRS frequency shift agains moisture content for various weights of soybean, reproduced from [5].*

Among the hundreds of applications of microwave aquametry, some of the more persistent examples include: fossil fuels, biofuels, turbine oil, assorted grains, green tea, tobacco, wood, concrete, veneer, paper, soil, snow, fabric and pharmaceutical powders and tablets (see above review articles for references). As well as listing literally thousands of publications, these authors all take pains to point out the significant advantages of microwave sensors, many of which apply to the whole array of MRS applications as well. A summary of the most widely cited follows in Table 2.1.

*Table 2.1: Advantages of MRS for aquametry*

| Advantage | Description |
|---|---|
| High contrast | Water has a microwave permittivity at least 10 times that of most dry materials, as well as a significantly higher loss factor. Such a large contrast means that its presence can often be singled out and estimated to a useful level of accuracy even with large variations in the underlying composition of the sample. Neither shorter nor longer wavelengths offer such a marked contrast and with so diverse a range of possible mixtures. |
| Non-destructive | Unlike the classic aquametry method of heating the sample for hours to drive off the moisture and measuring the resulting change in mass, or the accurate yet chemical Karl Fisher titration method [8] based on the reaction of water with a known concentration of iodine, microwaves do not affect the quality or composition of a sample in any way. |
| Noncontact | Because only the electric field of the sensor needs to reach the sample, a reasonable gap can be accommodated which enables samples to pass continuously past the sensor – ideal for production line automation. |
| Large penetration depth | Due to the centimetre wavelength of microwaves, they can be arranged to penetrate quite deeply within most non-metals without resorting to free-space propagation, allowing the resonator perturbation principle to be maintained, and avoiding the radiation loss and complexity of antennas. Because of this, the volumetric moisture content can be determined, rather than being limited to the surface as with other types of sensor. This avoids any assumptions about the homogeneity of moisture distribution; porous materials are generally drier near surface due to evaporation, for example. |
| Safe and non-ionizing | Another common way of estimating water content is by measuring density with a gamma ray densometer. However, such ionising radiation poses a potential health hazard, as it is carcinogenic. Any device seen to have the potential for such a risk to operators is increasingly unpalatable to industry, however small the threat posed may actually be. |
| Robust to extreme environmental conditions | Difficult environmental conditions include the presence of vapour, dust, surface contamination, as well as extremes of pressure and temperature. Popular with 'dirty' industrial production line settings, the environmental robustness offered by microwave sensor systems gives them an edge over more 'delicate' sensors such as infrared and capacitance cells, especially surface contamination as both are susceptible to this. However, this advantage is largely a result of careful product development, the flexibility of microwave resonator designs, and the lack of any need for moving parts or precision electronics in close proximity to the sensor, rather than any inherent advantage of microwaves. |

| Advantage | Description |
|---|---|
| Less sensitivity to ionic conductivity | This is often desirable as ionic conductivity is strongly temperature and composition dependent and tends to swamp lower frequency measurements. Since the effect of ionic conductivity is inversely proportional to frequency, by microwave frequencies its contribution to loss is much less significant, particularly in comparison to water, giving greater measurement robustness. |
| Can be independent of density | Various ways have been devised to make microwave moisture measurements independent of the density of the sample, which gives microwave sensors a big advantage over density-based aquametry methods when constant sample density cannot be guaranteed. |

Less often trumpeted (due perhaps in a large part to the inevitable bias of authors), are the disadvantages of microwave sensors. These are summarised in Table 2.2.

*Table 2.2: Disadvantages of MRS for aquametry*

| Disadvantage | Description |
|---|---|
| Expense of components | Sadly, microwave components (cables, connectors, amplifiers, electronics, etc.) are more expensive than RF components, and these are more expensive than low-frequency components. Although certain bands used for communications are cheaper due to mass demand, even these are still much more expensive than their low-frequency equivalents. For microwave sensors to be economic, they have to offer something that low frequency sensors do not – too often this essential practical point gets overlooked in the literature. |
| Large equipment dimensions | Despite several attempts to miniaturise the required equipment, many systems are still dependent on bench-top instruments to achieve the level of accuracy needed. Furthermore, because of conductor losses being essentially proportional to the surface area to volume ratio, microwave resonators cannot be made arbitrarily small without having an impractically low Q or undesirably high resonant frequency. In fact, most sensors need to be at least one-half wavelength long in one or more dimensions, leading to a difficult trade-off between the expense and unsuitability of working at higher frequencies or the large sensor size required at longer wavelengths. |
| Poor specificity | Because the 'dry' permittivity of a material is initially unknown and materials often show non-linear mixing rules with water, moisture estimation is inherently inaccurate. Although a variety of analytical mixing rules for composite dielectrics exists (for an overview see Sihvola [9]) many are based on tenuous assumptions such as non-interacting spherical particles or depend on unknown permittivity values, such as that of the 'bound water' layer that surrounds large solute molecules and ions due to hydrogen bonding. Thus in most cases the only way to provide an accurate reading is to fully characterise a range of material-water mixtures and use this as a calibration 'look-up'. However, this makes the sensor specific to a particular material and can introduce errors if the nature of the material or its environment changes in an unanticipated way. |
| Sensitivity to temperature | With water, the temperature coefficient of permittivity is particularly large across the whole microwave band; hence, precise control of temperature in aquametry experiments is particularly important. It is often difficult to determine the exact temperature of a sample inside a MRS, and a temperature differential between a sample 'reservoir' and the interior of the resonator is inevitable. Thus, even the introduction of a sample into a resonator can alter its temperature and result in a source of experimental error. |

| Disadvantage | Description |
|---|---|
| Sensitivity to environmental variation | As well as temperature, a multitude of environmental variables, such as surrounding air humidity, contact pressure, and the presence of other materials in the sample (contaminants, air, *etc.*) can have a significant effect on both the sample and the resonator itself. This generally implies that additional sensors or special procedures are required to correct for or minimise environmental variation in order to reduce measurement ambiguity – *e.g.,* misattributing a change in resonant frequency to sample composition when actually it is due to a change in temperature. This can easily undermine the advantages of a microwave sensor, particularly as the additional sensors and environmental controls can significantly increase the total system cost and complexity. |
| Limited spatial resolution | A less significant point since the advent of the microwave microscope (Section 2.3.4) has demonstrated that ultra sub-wavelength imaging is possible. However, what microwaves cannot do is achieve high spatial resolution as well as having large sample penetration. Shorter wavelengths, and particularly optical lasers, can be focussed on a small region of a sample over very large distances whereas microwaves cannot be 'focussed' to similar spatial resolutions without resorting to evanescent methods that have very limited depth penetration. |

Perhaps because of these disadvantages, lack of funding, or simply a lack of exposure, many of the proposed applications have seemingly not progressed beyond the lab. Although it is often hard to discern exactly how microwaves are being used in commercial devices, a number of companies can be found that currently manufacture MRS instruments for microwave aquametry. Sartorius-Omnimark [10] sells two types of microwave moisture analyser. One is based on the destructive method of microwave heating combined with sample lost-mass measurement that simply offers a faster way of drying a sample out – this type of microwave aquametry is sometimes confused with 'proper' microwave sensors. The other, however, is a genuine MRS for on-line moisture measurement aimed at process control that claims to measure moisture content from $0.1 - 60$ % with 0.08 % repeatability. The sensor appears to use a planar evanescent-mode resonant sensor not unlike the evanescent planar ring resonator developed by Reinhard Knöchel [11], proposed as part of a multisensory approach to determine the 'quality' of fish via a MRS method [12].

An example of a fully developed MRS system for microwave aquametry aimed at long-term structural health monitoring in civil engineering applications (*e.g.*, the detection of corrosion-related 'concrete-cancer' associated with reinforced concrete) was published recently by Sokoll and Jacob [13]. The miniaturised sensor (Figure 2.2) is based on a compact helical resonator measured in reflection mode. Unusually, the author goes into a lot of detail about the sensor electronics, both in this and several earlier papers [14], [15]. The heart of the measuring system consists of a single, the AD8302 manufactured by Analog Devices [16], which has a built-in phase and dual logarithmic detectors, claims a dynamic range of 60 dB, and is suitable for frequencies up to 2.7 GHz. The author makes a number of useful contributions to improve the accuracy of this miniature vector network analyser (VNA), including an electronically switched self-calibration system consisting of on-board offset short and offset open impedance standards. However, despite claiming a moisture level uncertainty of 0.7 %, the MRS system has a measurement period of 5 s and a poor SNR of the estimated resonator frequency response trace. Even so, this represents a very good example of the state of the art in low cost and low power MRS systems, having a peak power consumption of just 315 mW.

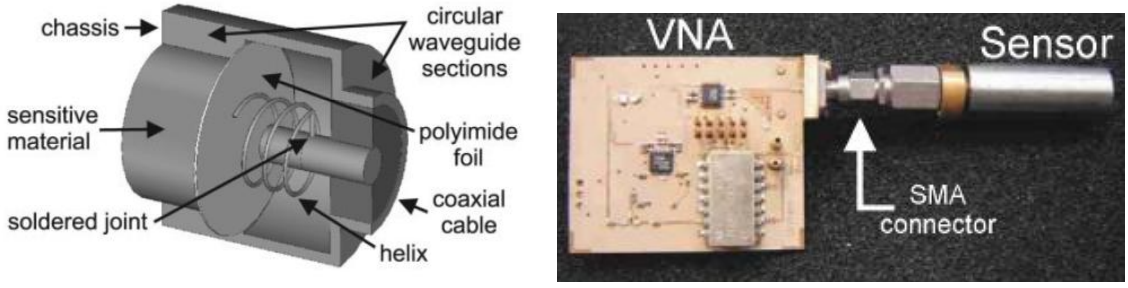



*Figure 2.2: Microwave resonant sensor (left) and compact measuring system (right) for the measurement of moisture content, reproduced from [13].*

The application of aquametry to structural heath monitoring leads to another application area for MRS: Non Destructive Evaluation/Testing/Monitoring (NDE, NDT, and NDM). More general than simply measuring water content, this application is all about measuring a change in material dielectric or sometimes

conductivity properties and relating that to a symptom of aging or damage. Examples include crack detection [17] and transformer oil aging [18].

Another closely related field to aquametry is that of compositional analysis. In fact aquametry is a special case of compositional analysis when the mixture has one part that is water. Densometry – the measure of density – falls in to this category too; in this case, one of the mixture components is air. Microwaves have traditionally been used to enable density and mass-flow measurement under continuous flow conditions [19], [20] where it would be impractical to measure mass directly and mechanical density meters would be unsuitable (in environments where there is too much vibration, for example).

Klein *et al.* recently developed a commercial liquid composition-based MRS system for use in diagnosing the contents of sealed liquid bottles [21]. The novel sensor uses a combination of a 100 MHz lumped element resonator used to measure conductivity, and a 2 GHz dielectric resonator with evanescent fields to measure complex permittivity. Combining these pieces of information is apparently sufficient to differentiate benign liquids from solvents or harmful chemicals, and was trialled at Prague Ruzyně airport to aid in passenger security checks [22].

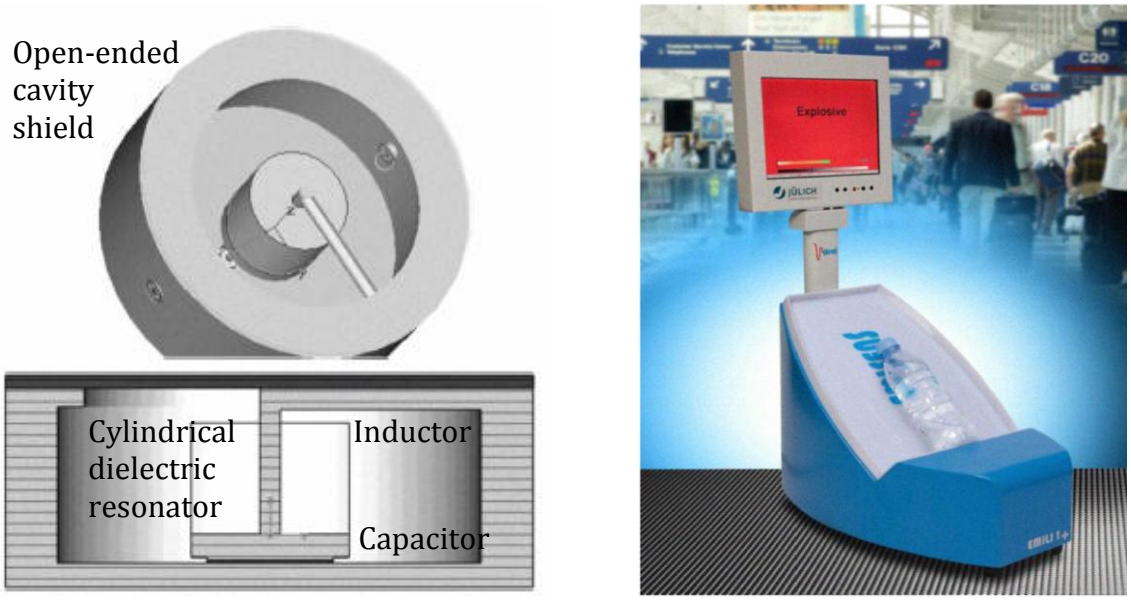



*Figure 2.3: Dual-mode resonant sensor with evanescent fields (left) and its application as a commercial prototype airport bottle scanner (left), reproduced from [21].*

One interesting early mention of this compositional measurement appears in a NASA technical brief in 1967 [23], where bubbles of hydrogen are detected in fuel lines (*i.e.,* measurement of liquid hydrogen density), presumably for use on the later Apollo mission rockets. Even over 40 years ago the authors state that "Microwave cavities have been used for many years to measure the densities of fluids and gasses". The 10 GHz sensor developed showed enhanced robustness to temperature and pressure variations over traditional capacitance sensors, and the thin coaxial fin design of the cavity end walls resulted in negligible flow disruption.

In a later technical note on the development of this MRS system [24], the authors also discuss one of two possible sensor readout schemes; the 'Resonant Frequency Tracking System', or REFTS and the 'Cavity Tuned Oscillator' or CTO system. The REFTS is interesting as it was an early attempt to create a high-speed measurement system as an alternative to the slow frequency counter method employed by the CTO; a basic self-oscillator method. It used frequency modulation to characterise the resonator, which was then used to produce a control voltage to keep the frequency locked on resonance. This in turn gave an analogue control

voltage from which the resonant frequency could be estimated (assuming the voltage-controlled oscillator (VCO) had a linear voltage-frequency characteristic). However, the authors seemingly failed to notice that the system would have been capable of full magnitude and phase characterisation if they had merely added a sweep generator to the VCO control signal.

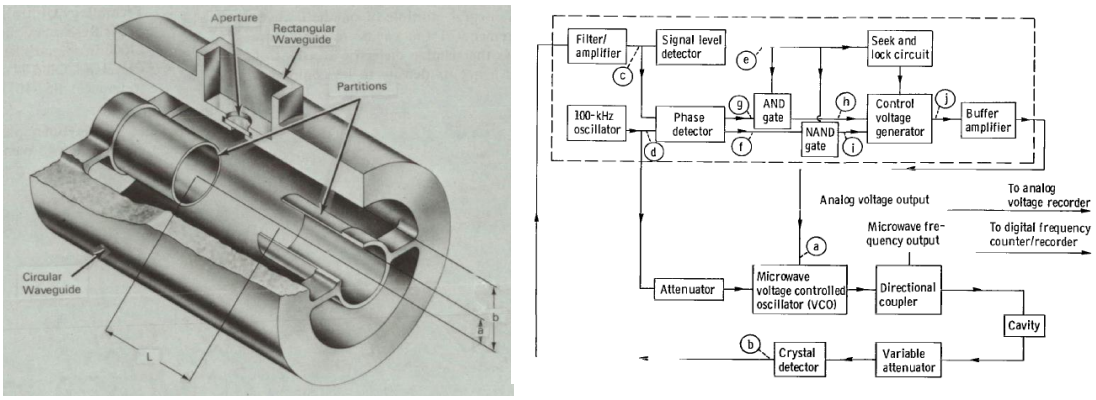



*Figure 2.4: In-line MRS (right) and readout system (left) developed at NASA between 1967-1971 for the measurement liquid hydrogen density in rocket engines [23].*

Measuring the density of very viscous flows is a widely commercialised application of microwave sensors, however only a handful of these use resonant technology. A typical example is in food production to ensure consistency of the end product; Toshiba manufacture a device which fits into a section of standard stainless steel pipe [25]. Other mixtures include the tricky multiphase gas, mud, water and assorted hydrocarbon concoction produced by oil and gas wells, where May *et al.* used a multisensor approach including a re-entrant cavity resonator [26]. MRSs have been applied to monitor such mixtures under fast-flowing conditions in pipelines for automatic process control in the extremely harsh environment of undersea well-head monitoring; such systems are commercialised by Roxar Flow Measurement AS, Norway [27]. This is a testament to how robust the MRS and its readout electronics can be engineered to be.

## 2.2 Scientific applications of MRS

Scientific applications related to solid-state physics are a strong, traditional application of MRS techniques. Primary among these is the study of superconductivity (see Section 2.3.1 for a brief introduction). Several scientific applications are reviewed in the third of the classic trilogy of articles on the cavity perturbation technique by Klein *et al.* [28], including the study of the temperature dependence of the conductivity of organic superconductor crystals such as tetramethyltetrathiafulvalene, the first 'unconventional' superconductor, discovered by Nobel laureate Klaus Bechgaard in 1979 [29]. Microwave resonant sensor techniques have also been instrumental in probing the properties of high temperature superconductors, both in single crystal form, *e.g.,* [30] and thin-film, *e.g.,* by the elegant method of patterning a coplanar resonator into it [31]. Regardless of the methods used, the study of superconductivity at microwave frequencies has a rich heritage and a number of practical applications in high frequency devices and systems, as reviewed in [32], [33], [34], and [35].

The general field of permittivity measurements has been reviewed several times by Jerzy Krupka [36], [37]. Resonant methods have been used to study the interesting microwave properties of nanomaterials, for example nanoparticles [38] and carbon nanotubes [39]. Ferroelectrics are a class of materials that have polarisation domains akin to ferromagnetic materials and show a nonlinear, hysteretic response with applied electric field; from their beginnings as merely an academic curiosity, they are now central to a number of high-tech applications [40]. Microwave resonator methods have been applied to their measurement in both bulk [41] thin film form [42], [43], including localised property measurement using a microwave microscope (see Section 2.3.4) [44]. This technique has also been applied to localised ferromagnetic resonance characterisation [45].

One of the few established uses for MRS in the medical field is in tumour detection, particularly in relation to breast cancer. Biological tissues have been extensively studied with microwaves over the years, and it was suggested as early as 1926 [46] that "malignant tumors have a greater polarizability than normal breast

tissues or benign tumors"; the higher dielectric constant of malignant tumour tissue being due largely to the enhanced blood flow required to support its aggressive growth. Microwave resonant sensors based on the evanescent coaxial probe model have been proposed for this purpose [47]. However, resonant techniques are in the minority as free-space imaging techniques and electrical impedance tomography, as well as high-resolution ultrasound and MRI scans, allow for improved 3D imaging and hence localisation capability that resonant sensors cannot match [48], [49].

MRSs have also been proposed for the measurement of temperature; the high Q and excellent stability of whispering gallery mode dielectric resonators, named after the resemblance of their circumferential EM field distribution to the acoustic resonance mode thought to be present in the whispering gallery of St. Paul's cathedral [50]. This lead the National Institute of Standards and Technology (NIST), Boulder, CO, USA to suggest their suitability as an alternative temperature reference standard, with an early prototype shown to have sub 10 mK reproducibility [51]. MRSs have also been suggested for pressure measurements with an elegant single-chip sensor [52]. However, it is not apparent how the use of an MRS in these non-specific sensor applications could ever be cost-effective when compared to its alternatives, restricting their use to a narrow range of ultra-precision experiments.

An interesting scientific use for microwave resonator measurements is in plasma research, *e.g.*, a simple hairpin resonator and basic swept power readout was found to give better accuracy than the expensive and widely-used Langmuir probe [53], [54]. Microwave resonators have also been applied to plasma generation; quite recently miniaturised resonators have been used to generate 'microplasmas' [55], [56] which can be used for analytical chemistry especially in 'Lab-on-a-Chip' applications (see Section 2.3.3). There still exists the apparently untapped potential for a self-measuring microplasma generating resonant sensor, based on the technology developed in this thesis for the self-monitoring microwave microreactor (Section 6.3), though quite how advantageous this would be is uncertain.

One curious and long-running series of microwave resonator based physics experiments relates to the search for (or more truthfully the progressive elimination of) a 'dark matter' candidate known as the axion [57]. The hypothetical axion, a weakly interacting yet very light particle, is postulated to interact with virtual microwave photons in a high-Q cavity under a large applied field (7.6 T in the experiment's latest incarnation), causing real microwave photons to be generated which would be detected close to resonance by an ultra-sensitive microwave receiver. Although microwave resonator characterisation is utilised solely for essential calibration purposes, the experiment undoubtedly represents the state-of-the-art in low noise microwave receiver technology and is probably the most sensitive microwave receiver ever built; the power expected to be produced by axion coupling is a mere $10^{-22}$ Watts [58], [59].

Since the 1950s [60], [61], and more recently [62], microwave resonators have been used to study the properties of gasses, and this is the only state of matter where microwave properties are accurately predicted and thoroughly understood at the quantum level. Unlike other states of matter, gasses exhibit energy level transition absorption lines that occur at microwave frequencies, corresponding to various rotational modes of dipolar species. Furthermore, the addition of a high voltage electric field leads to the separation in energy of E field aligned and E field anti-aligned quantum states (the level splitting is completely analogous to the principle behind magnetic resonance). Electron transitions between these split levels can absorb microwaves at a frequency proportional to the static electric 'bias' field – known as the Stark effect [63]. These various absorption lines can occur at frequencies throughout the microwave, millimetre and infrared spectra, with those of large molecules having high moments of inertia occurring at microwave frequencies in the centimetre region, 3 – 30 GHz. Collectively the measurement of these phenomena is known as microwave spectroscopy, see [64] and [65] for reviews.

The traditional method for measuring these spectra was to introduce a gas into an evacuated waveguide and measure the power transmission using a swept microwave source and power detector. After that resonant methods were

introduced, offering much higher sensitivities, but requiring a tuneable cavity (usually of Fabry–Pérot design) which inevitably led to long measurement times. In 1981, a major advancement came with the combination of a supersonic sample nozzle, which cools a beam of gas molecules down to ~ 1 K by its adiabatic expansion on egress from this nozzle when forced through it under high pressure. Another advance was the pulse-echo detection method, whereby the time-domain free induction decay (FID) of the molecular beam is detected by a sensitive receiver a fraction of a second after the application of a high power microwave pulse is applied. The FID signal is then Fourier-transformed in order to extract the absorption spectrum (which generally has absorption lines much narrower than the resonator bandwidth) [66]. The technique, known as Fourier transform microwave (FTMW) spectroscopy has been honed ever since, and the current advances of wideband arbitrary waveform generators (AWGs) and oscilloscopes have allowed for the direct (non-resonant) excitement of the molecular beam with a broadband 'chirp' signal pulse [67].

Whilst these recent developments look set to obsolete resonator methods in microwave spectroscopy, much as it has in the field of magnetic resonance, current research includes the miniaturisation of such instruments in order to make practical sensors; one interesting application is the detection of chemical warfare agents [68]. Here the attraction of microwave resonator sensors is renewed when low cost, small size and a restricted range of spectral measurement are required.

There are, of course, many more applications for microwave resonator measurements than there is space to mention. This is especially true in the pure scientific research applications discussed above, where the cost of instrumentation can be offset by the high precision and accuracy of the technique.

Yet the industrial applications discussed in Section 2.1, particularly moisture sensing, and composition sensing in general, demand relatively little from a microwave sensor. Physical variation and uncertainties in the constituents of a mixture, as well as assumptions made in the various mixing laws by which permittivity can be converted to composition, ultimately limit the accuracy of these

sensors. This means that there is no need for the sensor system to have better than about 0.1 % accuracy, achievable with relative ease using current technology. Furthermore, the requirement for simultaneous bandwidth and resonant frequency measurement is often inessential in these single-parameter industrial measurements. This all renders any readout technology improvements (and therefore any potential contribution from his work) somewhat redundant.

Despite this, there is still a potential need for further miniaturisation and lower power consumption in the industrial sensor area – particularly in the important fields of structural health and environmental monitoring. Whilst this suggests there is still room for improvement in the technology of microwave resonant sensor systems, this application space is definitely less demanding when it comes to requirements from the sensor electronics.

However, in Section 2.3 some rather more cutting-edge uses for MRSs will be introduced. These present a real demand for a better way of measuring resonators, not only in realtime but also with extreme accuracy. This is where the impact of improvements in the hardware and software of MRS readout systems can be most significant.

## 2.3 Dynamic applications

A central theme of this thesis is applying microwave resonant sensors in dynamic applications – where things are changing in time, perhaps quite rapidly. This section introduces some of these applications in more detail. Each of them could be improved by the advantages of faster sampling combined with high accuracy and integration. Many of these dynamic applications have been forced to use single frequency resonator characterisation to guarantee speed. This limits the dynamic range and accuracy of the resulting measurement, as well as making it sensitive to external sources of drift and other problems. Many of the applications mentioned here could therefore benefit from a high-speed spectral characterisation technique.

The main dynamic MRS applications that will be discussed are Kinetic Inductance Detectors (KIDs), sensors for microfluidic devices and scanning microwave microscope systems. Each is a good example of an application of microwave resonant sensors being part of a larger measurement system where speed, autonomy, accuracy, and overall system cost are important factors.

Research into KIDs highlighted some of the inadequacies of existing resonator readout methods, and an opportunity for improvement and a fresh approach in this area presented itself. The exacting requirements of KIDs include very high precision, high-speed sampling and massive readout parallelism to handle the desired numbers of individual resonator 'pixels'. Thus a new readout method was needed to retain the high precision of spectral resonator measurements – *i.e.*, where the response of the resonator is measured at more than one frequency – yet at the same time obtain each spectral estimate at a much greater speed than traditional measurement techniques. Finally, the architecture of the resulting system had to be scalable so that it could be duplicated a large number of times without becoming impractically expensive or prohibitively complex.

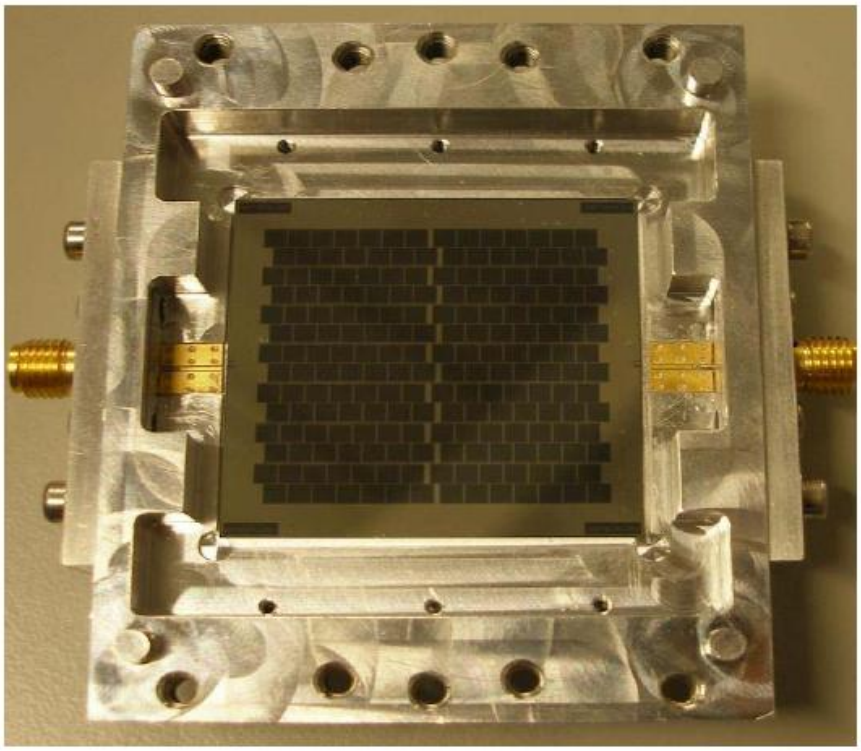


*Figure 2.5: Photograph of a 196 pixel Lumped Element Kinetic Inductance Detector (LEKID) array for mm wave detection designed at Cardiff University [69]. Each dark square is an individual microwave resonator consisting of a superconducting inductive meander and interdigital capacitor patterned by photolithography in a thin superconducting film.*

On-going research into microfluidics and continuous flow measurement at Cardiff University has previously shown [70] that MRSs had promise for diagnostic and compositional measurements of fluids. Here the objectives were the cost, footprint, and scalability of the readout system, as well as ability to sample at high speeds due to the flow rates and reaction times often encountered. A particular phenomenon of the laminar flow regime in these microsystems occurs when immiscible fluids are combined at a specially shaped junction. The fluids combine into a single stream of highly regular alternating fluid packets, known as segmented flow. This presents a particularly interesting challenge, and potentially an entirely new opportunity for real time microwave measurement – capturing each segment as it flows individually. Another opportunity was presented in the ability to regulate precisely microwave heating in microfluidic reactors by monitoring the characteristics of the microwave resonator used to deliver the

heating power. Since the small volumes and high flow rates present a dynamically changing environment, such a control system should be capable of fast sampling.

When it was first invented, the microwave microscope – a sub-wavelength imaging device based on evanescent field perturbation – was revolutionary. Since then microwave microscopes have been developed that have atomic level spatial resolution and can be integrated into conventional atomic force microscopes for automatic control of tip-sample separation. These systems have been used to make microwave frequency measurements of a host of fascinating nanoscale phenomena. Microwave microscopes generally raster scan line-by-line in order to build up a 2D image; however, this is very slow unless each 'pixel' can be acquired extremely fast. Not only is this tedious, it also creates a potential source of uncertainty if the sample is not truly held static throughout the scanning process. Another issue is synchronising the microwave readout with the tip-sample control; because this generally works in a modulated 'lock-in' mode, the microwave system should ideally sample at a high rate in order to extract this modulation from the microwave data. Furthermore, due to the extremely small relative energy perturbation that occurs at the tip-sample interface the changes in the host resonator can be minute. Thus, a microwave microscope readout system should also be capable of extreme precision.

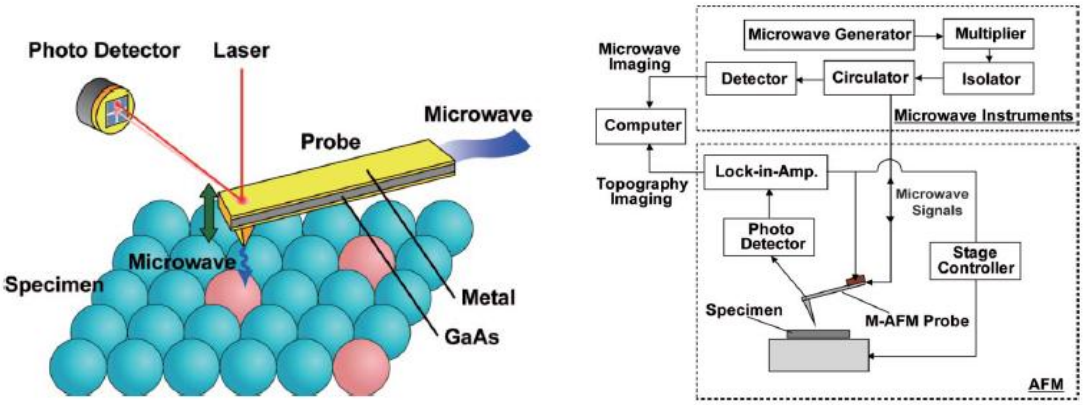



*Figure 2.6: Combined atomic force and microwave microscope (left) and control and measurement system (right), reproduced from [71].*

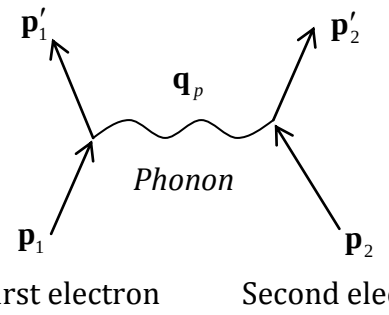
Because of the relevance of these applications to this thesis, there follows a brief discussion of these applications and their specifics, as well as how each in turn can be generalised into a problem that can be solved with a unified approach to readout and sensor design.

### 2.3.1 Introduction to microwave kinetic inductance detectors (MKIDs)

The loss-less super currents in conventional superconductors (which give rise to zero DC resistance and the exclusion of static magnetic fields – defining properties of superconductivity) are made possible by the formation of weakly-bound coherent pairs of electrons known as Cooper pairs, as explained in the theory of Bardeen, Cooper and Schrieffer [72], for which they shared the Nobel Prize in 1972 [73] – the second for John Bardeen, who was already a Nobel laureate for his part in the invention of the transistor. The principles behind the microwave kinetic inductance detector are intimately tied to the Bardeen, Cooper, Schrieffer, or BCS microscopic theory of superconductivity (as shown by the failure of the two-fluid model [74] to predict the correct sensitivity for MKIDs [75]). However, rather than attempt to describe this theory in all its quantum-mechanical rigor, for the purposes of this introduction there instead follows a 'classical' interpretation of the phenomenon which is altogether more lucid and helpful to the non-physicist.

The remarkable phenomenon that enables mutually repulsive electrons to be bound by an attractive force is the modulating effect each electron has on the metal ion lattice. This displacement of the metal ions from their equilibrium positions, caused by the attraction of a passing electron, creates a region of locally increased positive charge density that lags behind a moving electron, much as the wake of a ship modulates the surface of calm water. For very fast-moving electrons (those which are near the top of the Fermi distribution of kinetic energy, from the highest at the Fermi energy $\epsilon_F$ through a narrow band of width Δ), the lattice 'wake' stretches out into a long, narrow tube of positive potential and the displacement between its centre of charge and that of the electron is enough to create a small window of opportunity were a net positive attraction between two such fast-moving electrons can exist. The potential tube is too narrow to attract another electron except when it is approaching at the same speed from the opposite direction. Then, on passing, the electrons can become bound together; each electron finding an energetically favourable spot languishing in the trailing positive charge of the other, aligning nose-to-tail. This process has been likened to two children playing on a waterbed [76].

The bound electron pair then oscillate back and forth within a finite radius, known as the coherence length, and the lattice deformation becomes a bound displacement wave that travels with the pair. Waves of lattice deformation are quantised, like all waves, and have particle analogues called phonons, which get their name for their role in sound transmission. It is therefore said that Cooper pairs are bound together by an electron-phonon interaction, which follows from rigorous quantum mechanical analysis as well as the simplified classical picture described here [77]. In fact, there are also other types of lattice wave that can bind electron pairs together in this way, such as magnetic spin fluctuation waves [78] – thought to be responsible for superconductivity in some high temperature superconductors.




$\mathbf{p}_1' = \mathbf{p}_1 - \mathbf{q}_p$ First electron loses momentum…

$\mathbf{p}_2' = \mathbf{p}_2 + \mathbf{q}_p$ … which is transferred to the second via a (virtual) phonon.

$\Longrightarrow (\mathbf{p}_2' + \mathbf{p}_2') = (\mathbf{p}_2 + \mathbf{p}_2)$

Thus the net momentum of the pair is always conserved.

*Figure 2.7: Feynman diagram and description of the electron-phonon interaction.*

The radius of the Cooper pairs (*i.e.,* the mean average separation of the elections) is finite, on the order of $\hbar v_F/\pi\Delta$; the BCS coherence length ($v_F$ is the Fermi velocity, Δ is the energy gap). In conventional metallic superconductors this so large that many millions of other pairs can exist within the volume of one Cooper pair, and is several orders of magnitude greater than the lattice spacing. This causes long-range order effects such as the smearing out of superconductor-insulator transition regions, as the wavefunctions of overlapping pairs become coherent and pairs are no longer distinguishable as single entities. This coherency means that the ensemble of Cooper pairs moves as one through the lattice, without colliding or scattering – a superfluid in motion.

To be able to pass close enough for the initial binding to occur, the electrons must be of opposite spin, otherwise they would be excluded from having the same momentum (the Pauli exclusion principle). Thus, the pair of opposite-spin electrons make up a quasiparticle with zero net momentum and integer spin (a boson), since the two half-integer spins of the pair cancel. The distribution of energetically favourable 'orbits' for this two-electron quasiparticle is atom-like with the electrons occupying a high-order spherical S orbital having a banded structure of 'sweet spots' likened to an onion [79]. All the electrons within Δ of the Fermi energy will form these pairs, and these new bosons condense into a single, lower-energy zero-momentum quantum state. It is a phase change akin to freezing; for this reason it is said that the Cooper pairs form a thin frozen crust atop the Fermi sea of free electrons [80]. This creates a tiny energy gap Δ in the electronic

band structure of the superconductor that can easily be overcome by thermal excitation, confining the phenomenon to low temperatures.

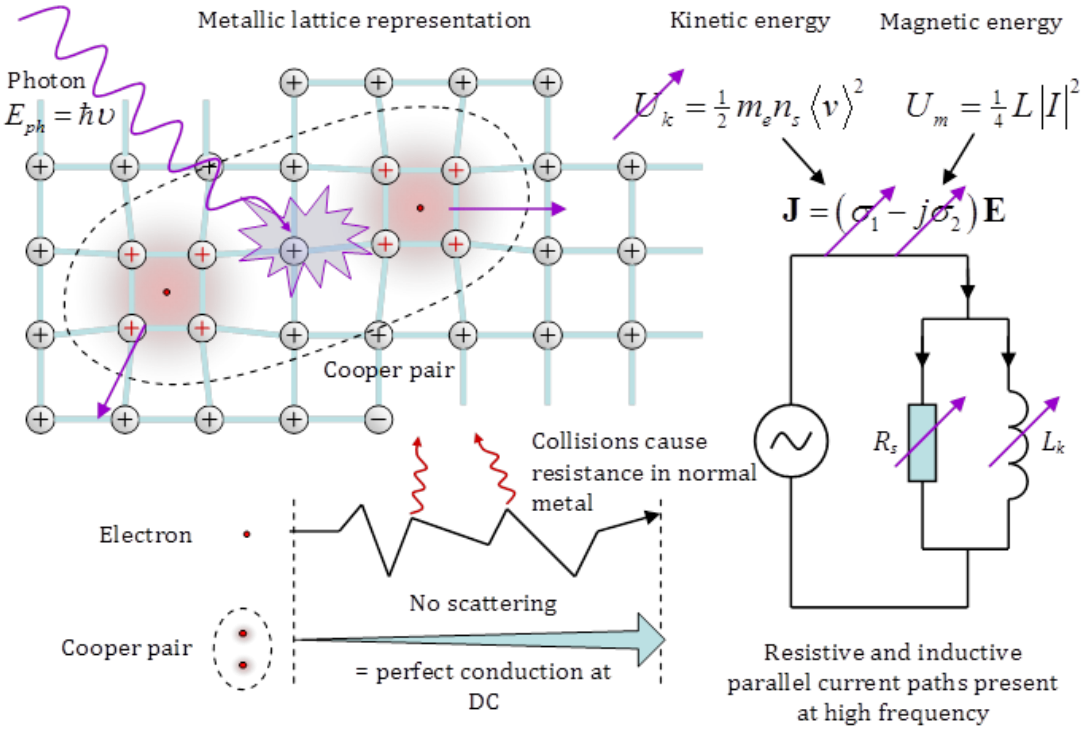
Cooper pairs travelling through the lattice under an applied field cannot scatter (the mechanism which causes resistance in normal metals), as each Cooper pair conserves momentum and thus cannot be scattered into a new energy state with a different momentum without supplying sufficient extra energy to break the pair up completely. This is because all the other available momentum states below the energy gap are already filled with normal electrons. Thus, it is the frozen crust of Cooper pairs and the protective energy gap it produces which prevents scattering and explains why superconductors have no resistance.

The superconductor energy gap is also central to the application of superconductors to photon detectors, such as KIDs. The small size of the energy gap allows low energy photons be absorbed. The photon energy breaks up one or more Cooper pairs, releasing ordinary free electrons (called quasiparticles, from Landau's theory of Fermi liquids that describes the free electrons in metals). These released quasiparticles reside in energy states above the gap – boosting the normal population that are there naturally due to thermal excitation.

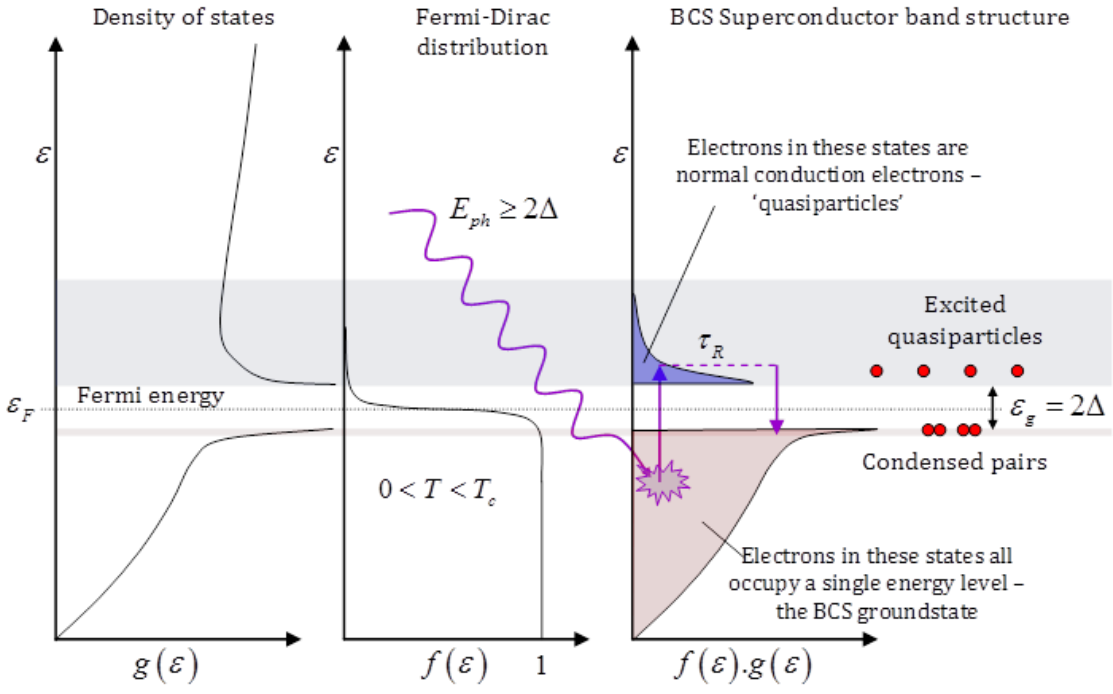
The excess quasiparticles then scatter about for some time, releasing the excess energy as phonons before reaching the right energy to recombine into new Cooper pairs and returning the population of Cooper pairs and quasiparticles to its normal equilibrium level. The population excess of excited electrons (and equivalent deficit of Cooper pairs) decays roughly exponentially after a single photon event with a constant called the recombination time, dependent on the efficiency at which the quasiparticle energy is scattered and how quickly the resulting phonons escape. This photoconductor-like model of superconductors is a direct consequence of an energy gap, and is applicable to most superconductors even if the nature of the gap (some superconductors have multiple energy gaps) and the mechanism of the Cooper pair coupling varies.

These two populations of charge carriers are both responsible for current flow under an applied field. However, the additional 'super current' flow of Cooper pairs has no resistance, and hence 'short circuits' any conduction by normal electrons. This results in the characteristic sudden transition to zero DC resistance as soon as the thermal energy is low enough to favour the formation of even one Cooper pair. However, when an alternating (AC) current is caused to flow, another effect comes into play. Due to the relatively long distances that a Cooper pair can exist and travel over without loss of momentum, they can amass a significant amount of energy as inertia when accelerated by an applied electric field – similar to a free electron in a vacuum tube. This kinetic energy may be viewed as equivalent magnetic energy storage – inductance – hence the name 'kinetic inductance'. The inductive and resistive current flows carried by the two conducting populations thus give rise to a complex conductivity, $\sigma \longrightarrow \sigma_n - j\sigma_{sc}$, and at microwave frequencies the surface impedance of a superconductor consists of a resistive and a significant reactive term: $Z_s \longrightarrow R_s + jX_s$.

At GHz frequencies, superconducting transmission line or wire thus has a small but significant loss at high frequencies, as well as an enhanced inductance: $L \longrightarrow L_{ext} + L_k$, where $L_{ext}$ is the standard geometric, or 'external' inductance and $L_k$ is the kinetic addition. This kinetic inductance fraction $L_k/L$ may be enhanced using specific geometries of conductor, namely thin films. A superconducting microwave transmission line or thin strip can thus be used as a detector – incoming radiation breaks Cooper pairs and creates excess normal electrons, which then increase $R_s$ and $X_s$. Note that fewer Cooper pairs actually store more kinetic energy for a given accelerating electric field – consider the analogy of inductors in parallel, one for each pair. Therefore, the kinetic inductance fraction is larger near to the superconducting transition temperature where there are fewer pairs. The change in surface impedance can thus be used to deduce the optical signal. This has been known for some time [81], but the insight to use a MRS to measure this change in surface impedance (see Figure 2.8 and Figure 2.9 for the principals of its operation) – and to operate instead at a low temperature to gain amplification by high Q factor – came more recently.

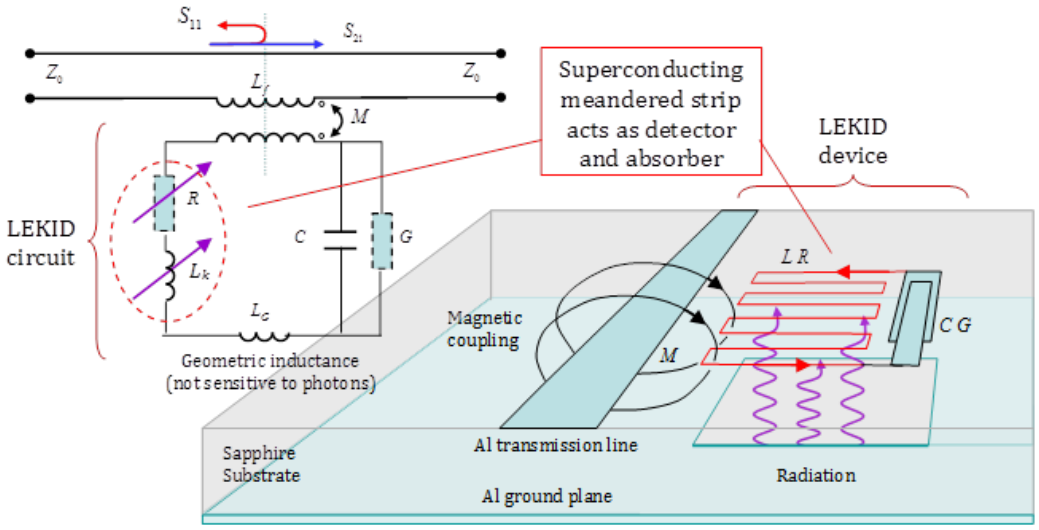



Photon breaks pairs, causing remaining Cooper pairs to accelerate more in the same electric field, increasing the stored energy and hence the kinetic inductance. In addition, the number of normal electrons increases, so the resistance also goes up.



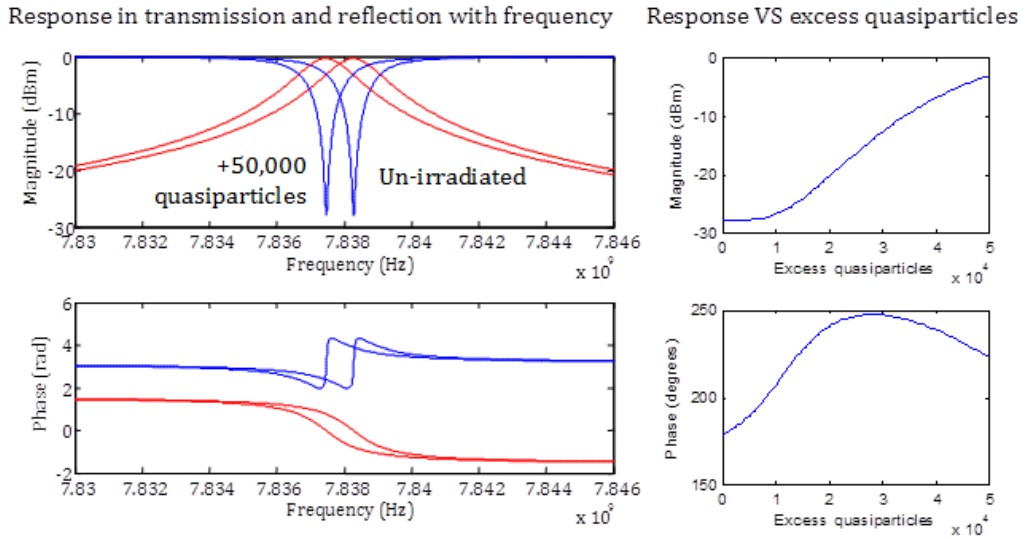

A photon of energy greater than the gap energy 2Δ will excite electrons from the BCS groundstate to higher energy quasiparticle states where they remain until they eventually relax back down again as they find energetically favourable pairings with other electrons. This characteristic time is known as the recombination time, $\tau_R$.

*Figure 2.8: Graphical summary of kinetic inductance detector principle.*




$$Z(s) = \frac{L_f s - M^2 (1/R + G) s^2 + \left(M^2 - L_f \left[L_k + L_G\right]\right) C s^3}{1 + \left[L_k + L_G\right](1/R + G)\, s + \left[L_k + L_G\right] C s^2}$$

Photons absorbed by the detector increase the resistance and inductance of the superconducting meandered strip. This alters the transfer function Z(s) of the LEKID circuit. The resonant frequency and quality factor will decrease, and the coupling will change. These changes can be determined by measuring the complex power transmission of microwave signals along the transmission line.




Analytical phase response estimate at resonant frequency

$$\frac{d\phi_{S_{21}}}{dN_{qp}} \approx \frac{2g}{1+g} \cdot \frac{Q_i \alpha \Gamma_1}{V_{eff}} \cdot \sqrt{\frac{2}{\pi}} \frac{\hbar\omega I_1(\zeta) - \left(2\Delta(0) + \hbar\omega\right) I_0(\zeta)}{N_0 \left(2\Delta(0) + k_B T\right) \sqrt{k_B T \Delta(0)}}$$

Measuring the response of the detector requires accurate, low noise measurements of signal amplitude and phase. Highest phase response is obtained by signals at the resonant frequency, and highest amplitude response is obtained by signals to each side of the resonant frequency, however better results could be obtained by measuring the response of the entire transfer function by looking at a number of frequencies.

*Figure 2.9: Graphical overview of resonant KID operation.*

### 2.3.2 Origin of the problem addressed in this thesis

The idea to use a microwave resonant sensor to amplify this change in superconductor surface impedance due to photon absorption was first proposed in 2002 by Day *et al*. from NASA's Jet Propulsion Laboratory (JPL) [82]. They showed that a microwave transmission line resonator made from a low temperature superconductor such as aluminium can have a very high sensitivity to photons, with a detector noise limit (known as its noise equivalent power, or NEP) as low as $1 \times 10^{-16}$ W/$\sqrt{\text{Hz}}$. The system they used to achieve this sensitivity relied on a homodyne, single frequency detection scheme with In-phase and Quadrature (IQ) readout, allowing for small changes in the magnitude and phase response of the resonator to be estimated at a single frequency. For a multiplexed detector, they needed one of these homodyne systems for each resonator. An analogue approach would be impractical for such a large number of channels on both cost and size grounds, so a digital solution was quickly sought. They based their multiplexed readout system around software defined radio (SDR) architecture (see Section 4.1 for brief background and discussion of SDR). This 'mostly digital' microwave resonator measurement system was the inspiration for my work, and the system architecture used by Mazin *et al*. [83] and later groups for KID measurements is the starting point for the developments of this thesis, both from a hardware and software perspective.

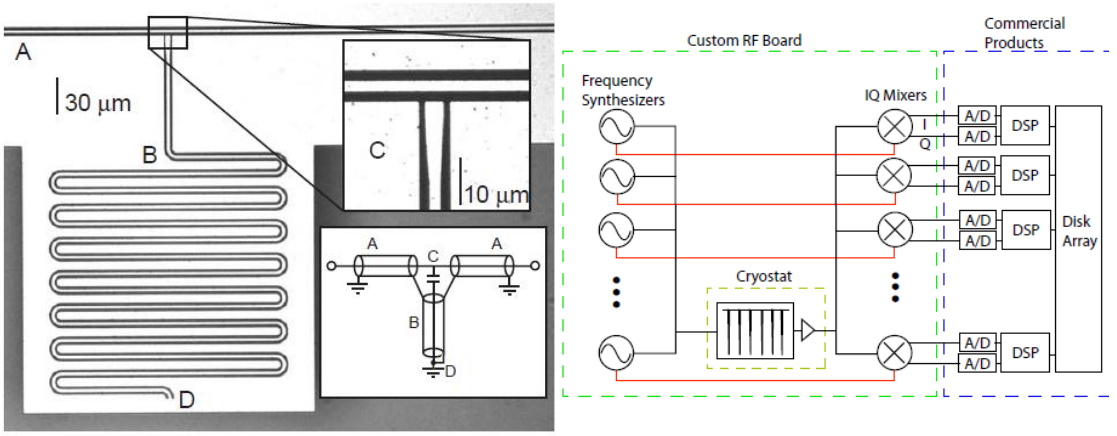



*Figure 2.10: Early aluminium transmission line MKID (left) and multiplexed homodyne readout system (right) developed by the Caltech team, from [83]*

During the subsequent study of KIDs, an unexpected source of phase noise was discovered [84] that resulted in a disappointingly high NEP. Phase noise is a common problem in resonator measurements; this source of error is usually caused by a combination of local oscillator frequency drift and instability of the resonator itself (*i.e.*, extrinsic and intrinsic phase noise). However, this noise source was different; it showed unusual power dependence. As the microwave excitation power was increased the noise level would decrease, eventually disappearing altogether, but only when the power was too high to be practical as it lead to overheating of the cryogenic refrigeration system.

This mystery snag was eventually blamed on defect-related, two level system (TLS) polarisation fluctuations in the dielectric of the resonator, creating a source of 'capacitance' noise due to the random fluctuations of TLS polarisation between the 'E field aligned' and 'E field anti-aligned' quantum states. This correctly predicted the observed power dependence of the noise: it was caused by the progressive 'saturation' of the TLSs. A high electric field amplitude reduces the probability that the polarisation of the TLS will flip randomly (*i.e.,* noisily) from one quantum state to the other, as the energy required to do so increases with E field and hence applied power [85] (extrinsic phase noise due to the RF source would not be expected to change with input power).

One proposed solution was to simply measure the change in insertion loss of the resonator on resonance – this 'amplitude direction' shows minimum sensitivity to phase noise, whilst still being slightly sensitive to the optical signal because resistive losses also increase with Cooper pair breakage [86]. In fact the sensitivity to photon flux has a vector angle of about −85°; in other words a large negative shift in resonant frequency (*i.e.,* phase) and a small shift in insertion loss (*i.e.,* amplitude). Thus, an optimal method was devised, which required measuring the IQ noise covariance matrix as a function of input signal frequency. In 2D, the contour of constant variance (*i.e.,* noise power) was found to be an extremely elongated ellipse (a Gaussian noise model was assumed), showing that the phase noise dominates amplitude noise. Computing the minor axis of this ellipse gives the vector of lowest noise, and the division of the signal and minor ellipse vectors can be used to estimate the direction of highest SNR, which is closely aligned to the amplitude, or I, axis (justifying the amplitude readout solution). However, the energy of the phase noise source is mainly concentrated at lower frequencies; thus signal inputs at higher frequencies have a different optimum vector direction for maximum SNR. By combining the in-phase and quadrature signals using an optimum frequency-dependent mixing matrix the best SNR can be found throughout the input signal frequency range. This approach was developed by Gao in his PhD thesis [86].

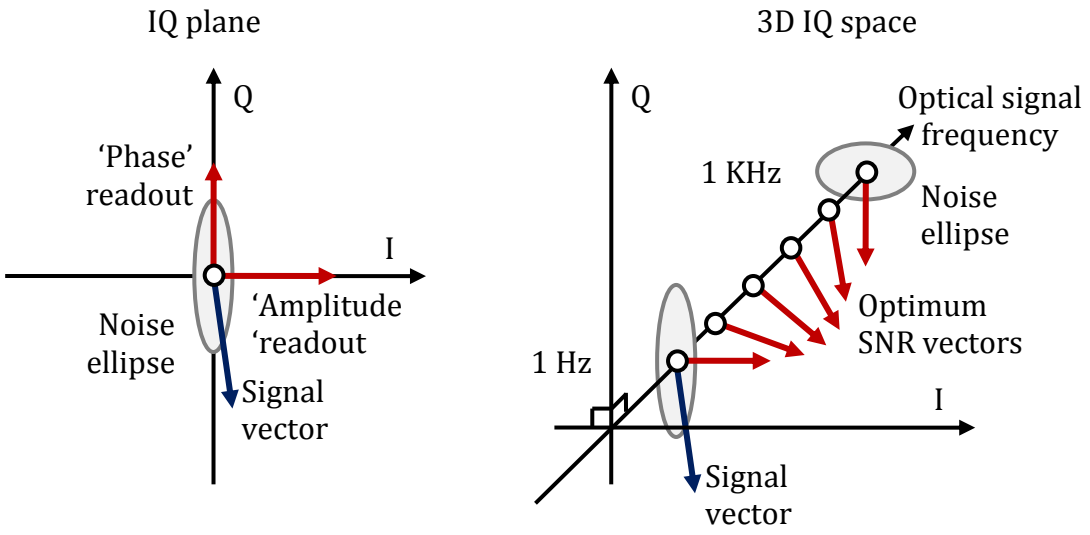



*Figure 2.11: Single-frequency IQ readout of a phase-noise corrupted signal. Simple 2D IQ plane representation (left) and 3D phase plane vs signal frequency space, used to account for frequency-dependent quadrature noise.*

However, this is still a single frequency readout method. The dynamic range is therefore inherently very limited; single frequency excitation cannot measure both small and large changes simultaneously (as an extreme, a large change in resonant frequency would result in the excitation signal being 'off resonance' – rendering the phase and amplitude measurements meaningless). However, the superior spectral characterisation method would require the ability to resolve the full resonator spectrum in real-time. Conventional methods are simply not up to the task. Thus, one of the main aims of this thesis was to devise an effective way of doing this without losing the low noise and high-speed advantages of the single frequency method. This system design and development is discussed later in the hardware section.

Even with spectral measurements the issue of optimum SNR vector is still present, and to make things more complicated this vector would vary not only with signal frequency but also with its position on the resonator spectral response curve. Thus, an optimum method would require the estimation of the best SNR vector over the whole resonator spectrum. Add an extra dimension to the existing 3D space and things start to become tricky to visualise. To handle this mathematically requires two spectral dimensions, one for 'signal frequency' and one for

'microwave frequency'. Therefore, another important aim of this thesis was to develop a comprehensive 'time-frequency' model of a microwave resonator sensor. This model is developed and discussed later in Section 3.4.

A different method to reduce phase noise issues, quite often used in astronomical detectors, would be to use signal chopping combined with a lock-in amplifier. This technique is potentially applicable to more microwave resonant methods than just KIDs, but here at least the method of 'chopping' the signal merely requires the telescope to 'jiggle' at a known frequency – effectively modulating the signal onto a known carrier [87]. This modulation of the signal can be used to place it far enough away from the $1/f$ distribution of phase noise to minimise its effects relative to unavoidable broadband noise sources (*i.e.,* Johnson noise or A/D converter quantisation noise). However, this method greatly reduces the bandwidth of signals that can be measured (it would be unsuitable for transient detection, for example). It is interesting to translate this technique to other MRS applications to see if, theoretically at least, a lower noise floor can be achieved through bypassing resonator phase noise problems. This alternative technique, and other ideas, will be studied in more detail in Chapter 4.

### 2.3.3 Sensors for microfluidic applications

Microfluidics is the technology of manipulating small amounts of fluids (gases, liquids and suspensions) on a small scale – typically when one or more dimensions of a flow system is smaller than 1 mm [88]. As things are made smaller, the balance of forces changes from those on the macro scale with which we have everyday experience. One of the key changes to take place in fluid capillary (small tube) systems is the increase in frictional forces (viscous forces), which resist and damp the motion of fluids, relative to inertial forces, which tend to keep the fluid in motion [89].

A very low viscosity to momentum ratio occurs, for instance, with the flow of air masses on everyday scales, and produces effects such as turbulence, which is difficult to model and causes unpredictable effects. However, within small cross-

section channels, a high drag to momentum ratio exists, even for gasses. In fact, fluids of different viscosities that on the macro scale behave very differently, such as glucose syrup and acetone, show exactly the same behaviour when pumped along microchannels. This is because a regime known as laminar flow exists, where flow occurs in straight, parallel lines and no rotation (vorticity) occurs when the fluid flows past an obstruction, for instance. Fluids in this regime behave very predictably, and the chaotic component of flow that causes turbulence is insignificant. This is because the energy stored in momentum is damped away before it can build up and feedback into the system, causing instability.

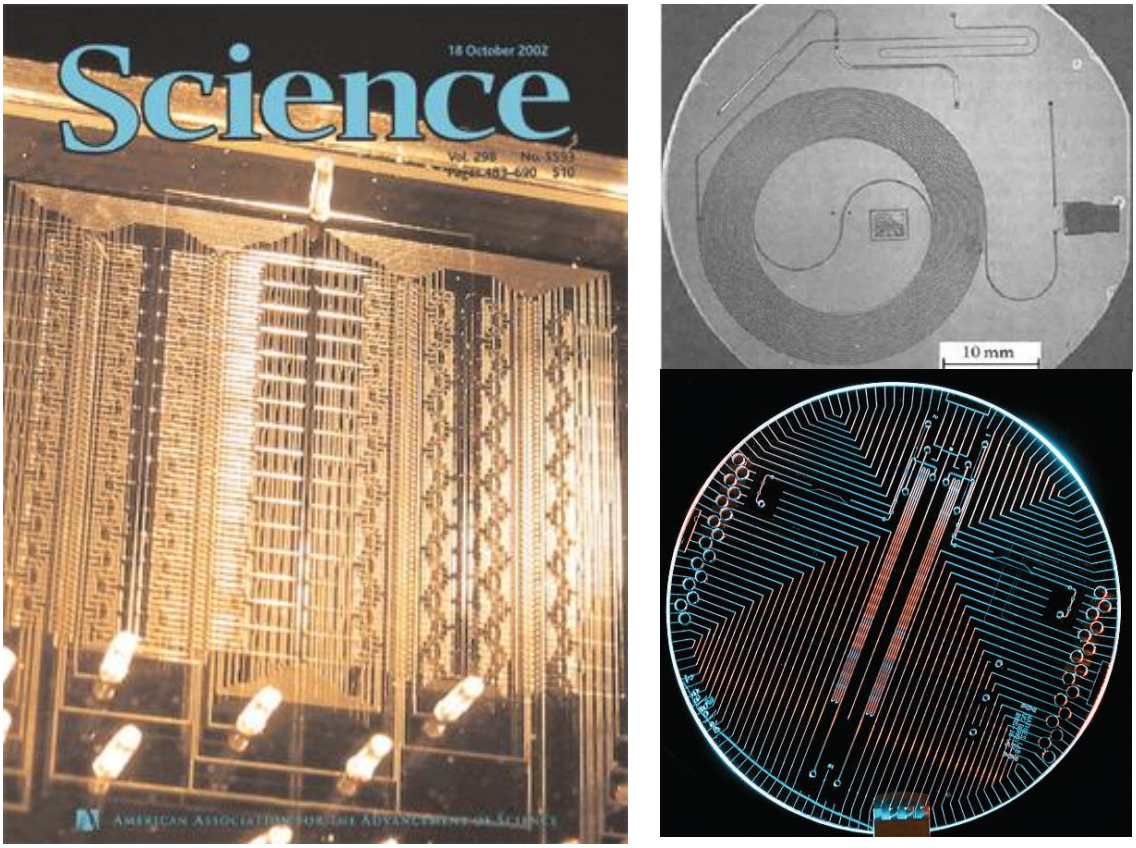



*Figure 2.12: Top right – the first 'lab on a chip', a gas chromatography system built by Terry in 1975 [90], left – a complex microfluidic device features on the cover of the journal Science, bottom right –a sophisticated microfluidic 'bioprocessor' for DNA sequencing [91].*

The predictability and regularity of laminar flow has given rise to a number of applications of microfluidics-based devices that would be either very inefficient or impossible to achieve on a larger scale. Furthermore, as the length scale (*i.e.,*

smallest channel dimension) of microfluidic devices is reduced, there are also a number of other beneficial effects. The larger relative surface energy tends to keep micro droplets more stable, and other surface phenomena such as selective adsorption, which is the basis for the separation of a chemical mixture by chromatography, become more significant. Electrically- and magnetically-induced forces also become significant enough to be useful. An example of this is dielectrophoresis, the induced movement of a fluid by a non-uniform electric field. This is the driving technology behind 'digital microfluidics' [92], the manipulation of discrete droplets of fluid in preference to continuous flows. These unique or enhanced phenomena, with more than a little hype [93], have helped make microfluidics the technology of choice when precision control, measurement and manipulation of fluids is required.

One particular branch of microfluidics is the Lab-on-a-Chip (LoC), also known as a Micro Total Analysis System (uTAS) [94]. This is a paradigm where different chemical analysis and synthesis operations are combined into a single microsystem, where fluid analytes are handled and contained using microfluidic techniques, and the ancillary electronics and systems are to be miniaturised and integrated as far as possible. The ultimate aim is therefore to create a miniaturised laboratory complete for a particular task. This could be investigating cells to check for diseases or cancer, performing assays on bodily fluids to diagnose disease and monitor health, particularly in the home or in places lacking access to proper healthcare, to check the efficacy of drugs without in vivo testing, environmental monitoring in remote locations to investigate pollutants, *etc*. Such systems rely heavily on sensors, especially in terms of the definition used earlier, *i.e.,* where a low level of manual intervention and a high level of intelligence, integration and robustness are required.

A distinct but highly overlapping branch of microfluidics can be termed "industrial microfluidics", or micro process engineering. Rather than being aimed at reproducing analytical laboratories on a chip, this technology is aimed at creating miniature chemical plants and factories (known as microreactors) aimed at synthesising chemicals rather than analysing them. Driven by the potential

increases in efficiency and performance (*e.g.*, purity of end product) attainable due to the highly controlled environment of micro flow, especially rapid thermal cycling and short diffusion lengths, this technology is being considered for on-demand synthesis, perhaps in remote locations, and to enhance yield of difficult reactions or those requiring complex steps [95]. This interest is helped in a large way by 'green' considerations such as the reduction of hazardous waste, reduced reagent consumption, containment of toxic, carcinogenic or explosive intermediate reaction products and the reduction in energy usage, all of which microfluidic technology claims to offer.

What does microwave sensor technology have to offer these applications of microfluidics? As a sensor, most established microwave resonant sensor types are not suitable for LoC applications. Their physical bulk and extensive supporting hardware, as well as the lack of the autonomy and robustness that is also required – either in the sensor or in its electronics – has meant that the integration of microwave resonant sensors with microfluidic technology has not advanced beyond a few proof-of-principle academic demonstrations. Examples include the substrate integrated waveguide sensor, which has potential for fluidic channel integration due to its compact size and reasonable $Q$ factor (700 at 8 GHz) [96], [97], [98]. Shaforost *et. al* [99],[100] made dielectric resonator based sensors at both microwave and millimetre wave exploiting the very high Q factor whispering gallery resonant modes to detect a nanolitre volumes of liquid. However, they did not actually address the microfluidic integration problem, as the liquid had still to be pipetted into the sensor manually. Whilst the whispering gallery mode resonator is already a powerful and developed technique for biosensing at optical wavelengths [101], as a microwave sensor it will always tend to be impractically large and expensive. Previous work at Cardiff University has also made inroads into this problem, culminating in a highly sensitive sapphire dielectric resonator with fully integrated circumferential microfluidic channel and active sensing volume of 56 nanolitres [70].

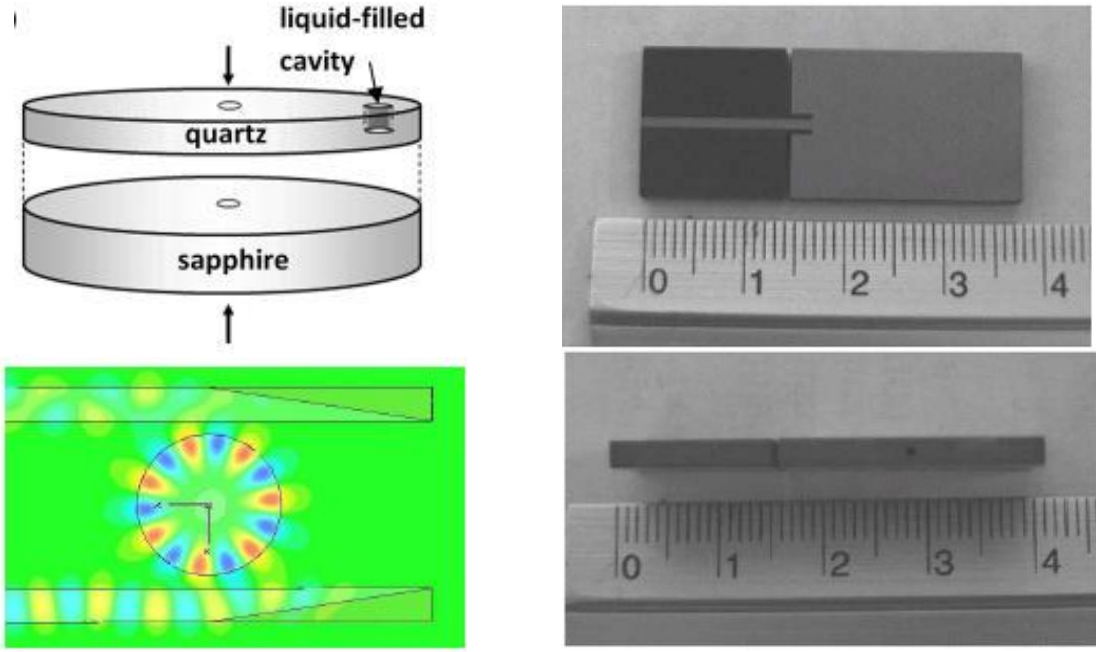
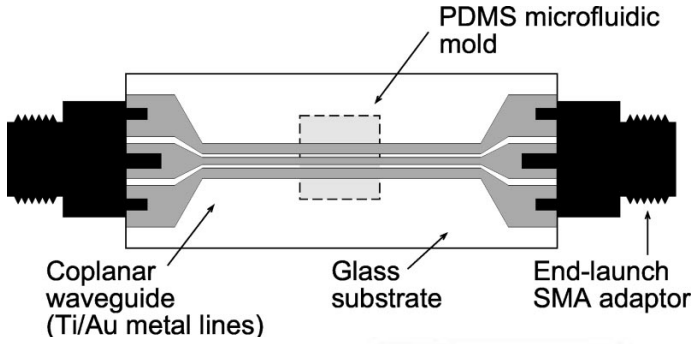



*Figure 2.13: Microwell whispering gallery mode resonant sensor with simulated electric field distribution with coupling waveguides (left), [100] and a substrate integrated waveguide cavity resonant sensor with hole for fluid capillary (right) [96].*

Basing miniature resonant microwave sensors on planar lithography or micro electromechanical systems (MEMS) fabrication technology is a distinct possibility – this technology has been used to integrate broadband waveguide structures, typically coplanar waveguide (CPW), with microfluidic devices on several occasions. Initially proposed by Facer *et al*., [102] the technique was gradually refined and improved in accuracy by the group at National Institute of Standards and Technology (NIST, based in Boulder, CO, USA) [103]. Others have also reproduced this work and applied it to 'biosensing' within microfluidic channels [104]. However, these broadband measurements require an expensive VNA and laborious calibration routines. Also, as discussed in the Theory chapter, non-resonant methods have much lower sensitivity than resonant methods, thus requiring larger liquid volumes for a given sensing precision.

(a)




(b)

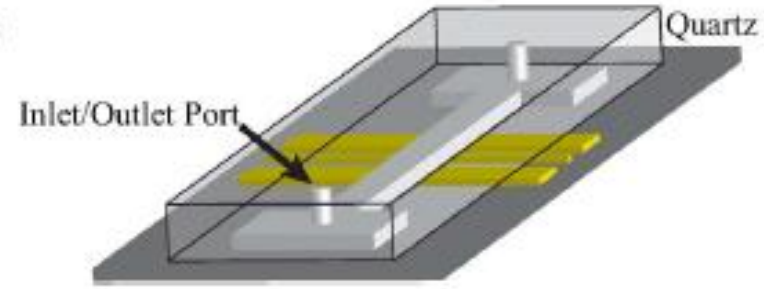



*Figure 2.14: Two implementations of a CPW with integrated microfluidic channel for broadband microwave permittivity measurements. Shown in profile from reference [102] (a) and in 3D from reference [103] (b).*

Moving from this low sensitivity broadband method to a resonant sensor could be as simple as creating two capacitive coupling gaps in these transmission line structures, as was done crudely over 10 years ago in an attempt to make a water-ethanol-glucose compositional sensor based on a microstrip line resonator [105]. Others have exploited variations on a simple transmission line, realising that a higher sensitivity can be achieved using the higher filling factor of lumped element components. Adapting the MEMS technology used to make various types of 'biosensors', [106] transmission lines periodically loaded with lumped elements (distributed MEMS transmission lines – DMTL) were fabricated in order to create propagation conditions sensitive to the localised dielectric properties of an introduced fluid [107]. It seems a mystery, however, why a single miniature lumped element microwave resonator would not be better than this overly complex approach. Converting a miniature lumped element resonator such as those in [108] to a resonant sensor using this MEMS fabrication technology could result in an extremely high filling factor, highly miniaturised sensor. However, it is possible that the quality factor would be too low for it to be possible to measure its spectral response accurately.

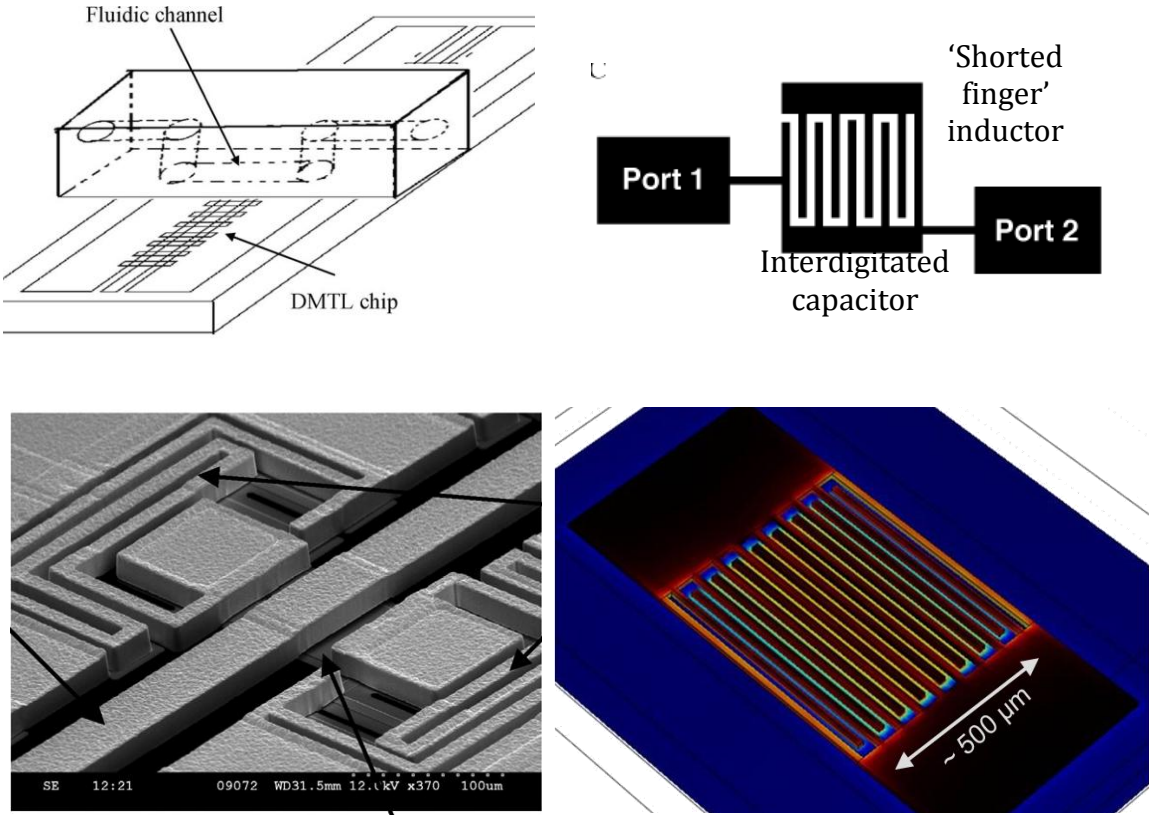



*Figure 2.15: Clockwise from top left: schematic of microfluidic integration of MEMS based transmission line sensor [107], compact lumped-element microwave resonator design [108], scanning electron micrograph of the active region of the MEMS transmission line device [107], and simulation of a hypothetical 7 GHz lumped element resonant sensor in a microfluidic channel. In the last figure, the electric field magnitude is shaded dark blue – red and current density black – orange.*

However, it is in the context of micro process engineering – the job of making new chemicals, drugs, and materials with microfluidics [91] – that I believe microwave resonant sensors have the biggest potential. Already quite extensively used in some industrial processes as discussed earlier, microwave heating and to a lesser extent microwave sensing is a mature technology in industries such as lumber, paper, food, and agriculture. Microwave-assisted synthesis is an efficient and powerful technique that has been around for several decades [109]. Attempts by others to integrate microwave heating with continuous flow systems have enjoyed some success [110]. The power of microwave resonators to both heat and act as a

sensor for continuous flow systems simultaneously, something recently demonstrated for the first time at Cardiff University using the resonant sensor real-time measurement techniques developed in this thesis, could be a promising technology to support the aims of the industrial microfluidics community.

One demanding but potentially significant application of microwave resonator technology – both sensing and heating – in microfluidics is in droplet flow systems. One interesting and useful phenomenon that occurs due to the relative enhancement of surface forces and the laminar flow regime is that of the segmentation of a multiphase flow, reviewed by Günther and Jensen [111]. Two immiscible fluids combined at a junction (or one injected into another) will form very regular droplets under certain circumstances, or elongated 'slugs' when in a channel geometry of dimensions too small to support a more energetically favourable sphere. Droplet generation is not only regular but can be very fast – as high as 10,000 droplets per second [112]. Including the increasingly important field of digital microfluidics – where droplets are manipulated on an individual basis – and an opportunity for a sensor capable of measuring and possibly also heating individual droplets presents itself.

Apart from miniature integrated resonators, suitable geometries of resonator already exist, and are based on the concentration of fields to a point using a sharpened probe coupled to a conventional cavity or dielectric resonator – a device identical to a 'microwave microscope' but without the need for a translation stage [113]. Such a device must be capable of sampling very fast – this is the key to it offering an advantage over competitive techniques – and it must be precise. One excellent technology for sensing and actuating individual droplets, particles and even cells is low frequency (up to several MHz) excitation by microfabricated electrodes, *e.g.,* [114]. Microwaves must offer a unique level of precision or function to be competitive with this well-established technology.

### 2.3.4 Introduction to scanning microwave microscopes

Most scanning microwave microscopes are a type of MRS, although some are broadband (non-resonant). Typically, they consist of a macroscopic resonator body, often coaxial but increasingly dielectric, that has a protrusion, or 'probe' with a sharp tip designed to create a highly localised field region which is then perturbed by a sample. The sample is scanned back and forth underneath the probe in order to build up an 'image' derived from the electromagnetic interaction of the sample with the localised field at the tip. The technique is a powerful tool for investigating the fundamental properties of micro- and nano-structured materials, and has become highly sophisticated over more than two decades of continuous development.

The classical resolution limit – the minimum spacing that two point-like objects need to have in order to be distinguishable by a wave – was previous thought to be limited to around $\lambda/2$ (the Abbe diffraction limit). However, it was found that evanescent, or 'near field' waves can overcome this – these rapidly decaying waves occur naturally when a propagating wave meets a discontinuity beyond the cut-off for propagation, such as at the edge of an optical fibre, outside of a sub-wavelength hole or at the end of a transmission line. Microwave microscopes for dielectric sensing (the most common type) exploit the evanescent electric field concentrated between a small radius conductor and a distant ground plane in order to exceed massively the Abbe limit. The idea for near field imaging was first proposed in 1928 by Synge [115]. Early practical demonstrations are Soohoo in 1962 [116] and Frait in 1959 [117], who achieved sub-wavelength imaging in microwave magnetic measurements. Ash and Nichols [118] are usually credited with the invention of the microwave microscope for dielectric measurements as it generally understood today. The technique was later extended to visible light by Pohl *et al.* [119] and Lewis [120]. The sharpened metallic tip design which has become standard was proposed by Wei *et al.* [121]. Later improvements in resolution came with tapered microfabricated waveguides, such as parallel plate [122]. Many attempts were made to integrate microwave microscopes with other types of scanning probe microscopy, such as Atomic Force Microscopy (AFM). Tabib-Azar *et al.,* attempted

to microfabricate a tiny coaxial aperture on a conventional AFM tip [123], but this early attempt was non-resonant and used a basic power detector readout.

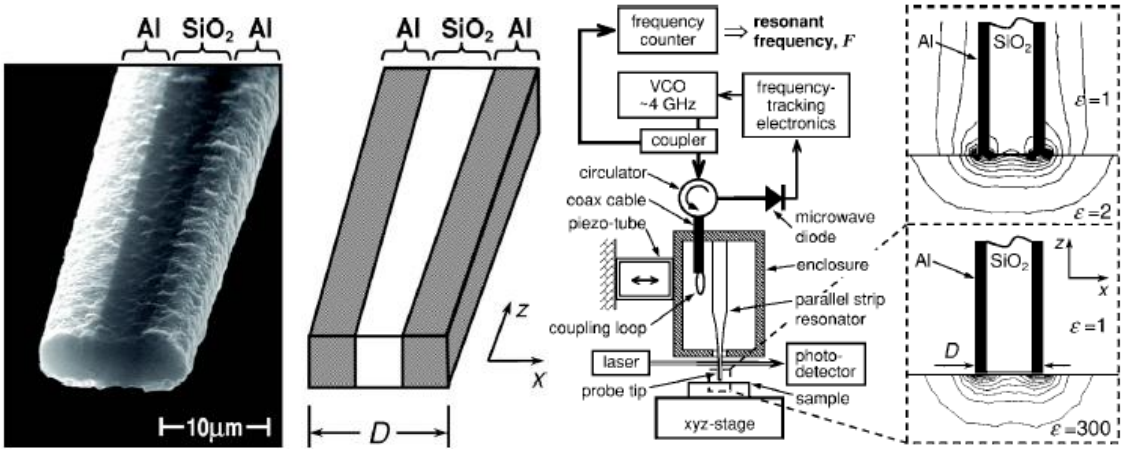



*Figure 2.16: Microfabricated tapered parallel plate probe (left) and primitive power readout along with basic finite element method simulations (right) in an attempt to deduce an inversion model [122].*

Despite these advances, there were difficulties with inversion that meant that the majority of authors made no attempt to convert resonant frequency and bandwidth data into material parameters. A valiant attempt to use a combined FEM and semi-analytical inversion model [124] showed that sample thickness was a significant source of uncertainty as it changed the effective stored energy of the resonator. The difficult convolution of the effects of variable tip-sample distance during scanning and varying spatial dielectric property was partially resolved with the introduction of tip-sample separation control mechanisms; however, robust permittivity inversion remains an issue that lacks a fully satisfactory solution to this day [125].

One method to control the tip-sample separation is combining conventional Scanning Tunnelling Microscopy (STM) with the microwave microscope. The STM feedback mechanism then keeps the tip-sample separation constant by maintaining a constant tunnelling current. Recently this method was used to produce images of individual gold atoms at a 2.5 GHz [126], and individual Gallium and Arsenide atoms [71]. Another method, which has the advantage of not requiring conducting samples, is based on shear force feedback using a

tangentially excited mechanical tuning fork attached to the resonant probe tip. The resonant frequency and Q factor of this tuning fork decrease when the separation approaches single nanometres, a phenomenon thought to be related to enhanced squeezed-film damping, capillary forces, or van der Waals forces; this change can then be used as a feedback signal to maintain a particular nm separation. First applied to near field optical microscopy [127], the technique was adopted by the collaboration lead by Kiejin Lee and Barry Friedman who applied it to glucose biosensing, among other applications, [113], [128], [129], one of these being the microfluidics-relevant microwave microscope based 'biosensor' which can be used in a non-scanning arrangement to probe small fluidic systems.

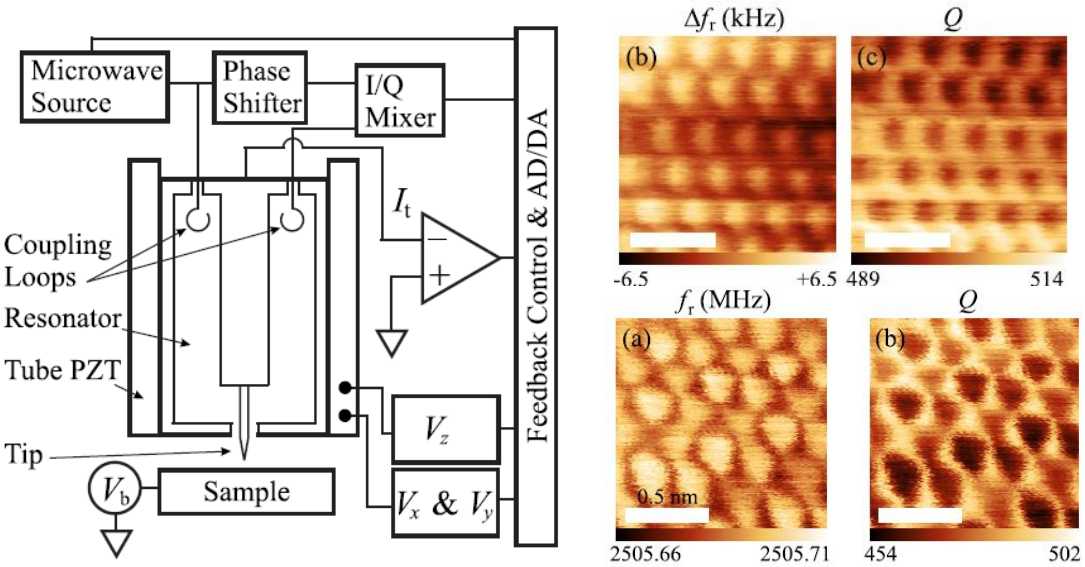



*Figure 2.17: Schematic of a combined Scanning Tunnelling Microscopy (STM) and Near Field Microwave Microscopy (NFMM) system, using quadrature readout and an analogue tunnelling current lock-in system (left). Images acquired by the STM-NFMM (left) showing microwave measurements of a gold surface with individual atoms resolved. The lower set of images shows the effect of deactivating the STM tip-sample separation control [126].*

A system combining microwave microscopy with Atomic Force Microscopy (AFM) has been commercialised by Agilent Technologies [130]. The device is unusual in that it employs the customary half wavelength coaxial resonator with sharpened tip, but as an impedance transformer that is not decoupled from the line, apparently to give a sensitivity boost whilst still allowing absolute measurements

of capacitance. This is useful for its prime intended application in the measurement of local semiconductor carrier density. This material property can then be estimated via derived dC/dV characteristics obtained using amplitude modulation of the excitation voltage and lock-in detection methods [131].

Other applications to benefit from the microwave microscope include semiconductor research, [132] ferroelectrics, [133], [134] transparent conducting oxides (TCOs) [135], organic semiconductors [136], and self-assembled monolayers [137]. The recent developments in atomic-scale microwave imaging have led to their proposed use in the study of quantum dots, low dimensional systems that show fascinating fundamental properties owing to their quantised energy levels [138]. The authors are apparently unaware of advances in atomic resolution microwave microscopes, but despite their primitive experimental setup (consisting of a microstrip resonator and diode detector readout) claim an impressive capacitance sensitivity of $10^{-19}\ F/\sqrt{Hz}$ – enough to measure the dC/dV characteristics of a single quantum dot.

Another rich scientific application is the measurement of localised superconductor properties, including effects at the grain boundaries of High Temperature Superconductors (HTS) [139]. Notably from the instrumentation point of view, as part of this work it was realised that broadband measurements could be done whilst still retaining high sensitivity through the use of a multimodal resonator [140].

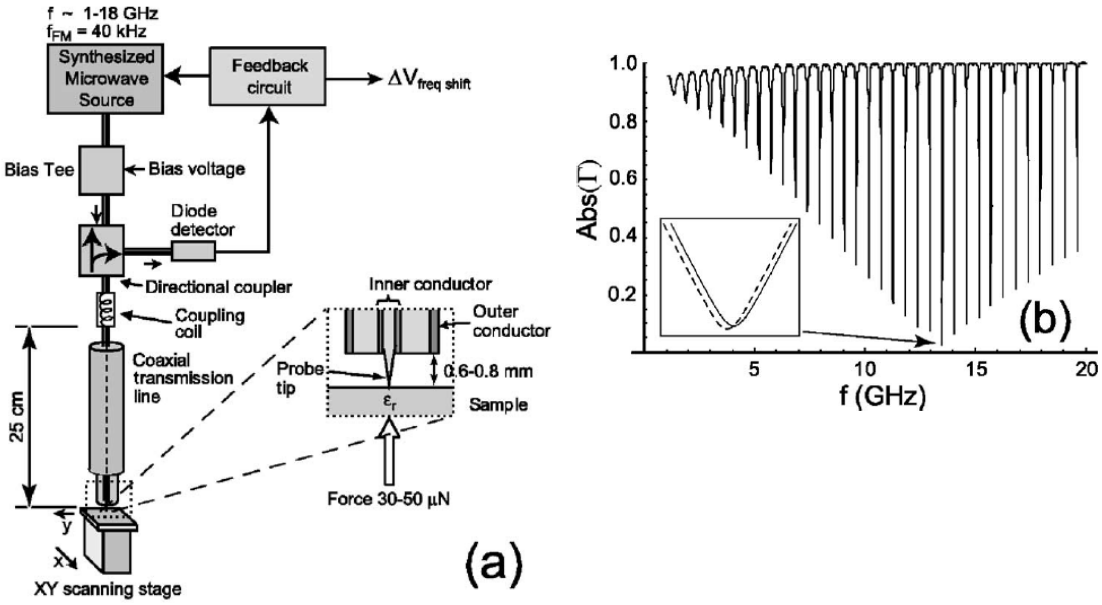



*Figure 2.18: Multimode resonator used as a broadband measuring device in a scanning microwave microscope stage used in the study of high temperature superconductors (left). The plot showing the large number of absorption resonances (right) highlights a central issue in multimode resonator – the problem of non-uniform coupling across the band [140].*

Most authors pay little attention to improving the microwave resonator readout technique, largely choosing to emulate a conventional single-frequency readout or using a commercial VNA. However, the improvement of resonator readout technology, as well as a detailed analytical inversion model, is discussed in a patent by Xiang *et al.* [141], providing a useful reference despite the inevitable patent bias. According to the author, of the methods proposed for microwave microscope readout the analogue phase-locked loop method is limited by stability, conventional swept measurements are limited to 20 Hz, yet swept measurements with a fast Direct Digital Synthesis (DDS) source can apparently be up to 10 KHz. Because of this, in their patent a fixed-frequency DDS method with IQ readout is suggested, which incidentally is the same method as used for the latest kinetic inductance detector systems. Data rates of 100 KHz – 1 MHz are claimed, but the authors acknowledge the severely restricted dynamic range offered by a single-frequency method.

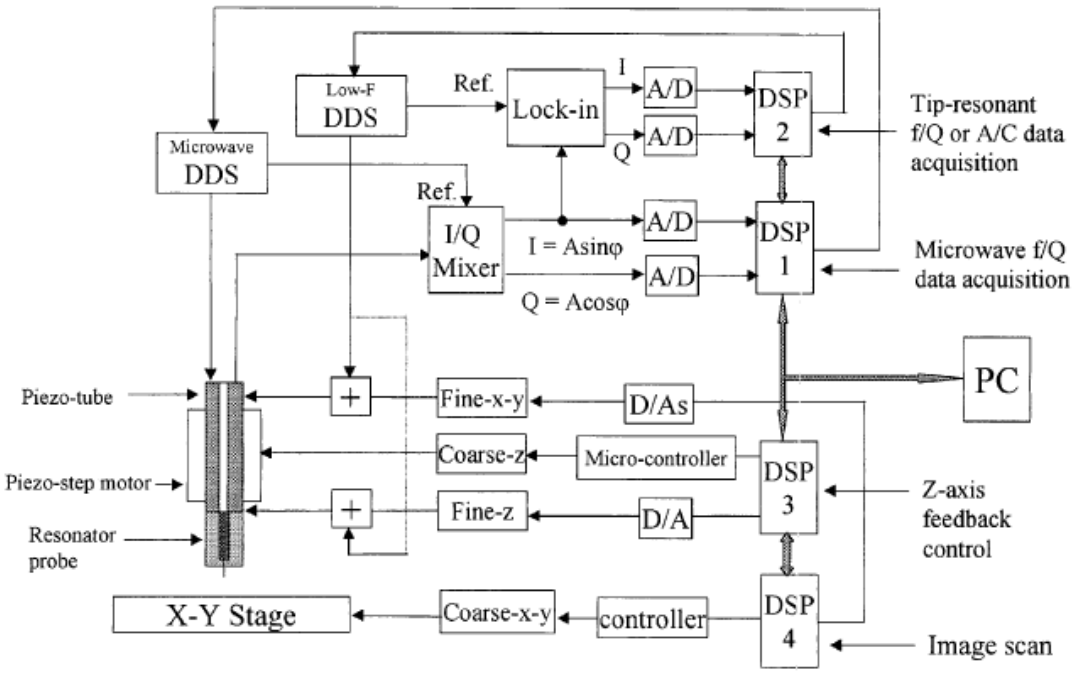



*Figure 2.19: Control system for a microwave microscope based on microwave DDS and single frequency quadrature readout [141].*

This chapter has been a summary of the applications of microwave resonant sensors, concentrating on demanding, high-performance measurement. In the next chapter the theoretical background to their operation will be given, and models are derived that can be used to understand and correct for their imperfections. A comprehensive theoretical treatment is important to ensure accurate and reliable operation at the fastest possible readout rates. This will establish that MRS can offer the performance necessary to match some of the cutting edge applications discussed here.

## 2.4 Chapter references


[1] U. Kaatze and C. Hübner, "Electromagnetic techniques for moisture content determination of materials," *Measurement Science and Technology*, vol. 21, no. 8, 2010.

[2] E. Nyfors, "Industrial microwave sensors—A review," *Subsurface Sensing Technologies and Applications*, vol. 1, no. 1, pp. 23–43, 2000.

[3] E. Nyfors and P. Vainikainen, *Industrial Microwave Sensors*. Artech House, 1989.

[4] S. I. Okamura, "Microwave technology for moisture measurement," *Subsurface Sensing Technologies and Applications*, vol. 1, no. 2, pp. 205–227, 2000.

[5] A. W. Kraszewski, "Microwave aquametry-needs and perspectives," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 39, no. 5, pp. 828–835, 1991.

[6] A. Kraszewski, *Microwave aquametry*. IEEE Press, 1996.

[7] K. Kupfer, *Electromagnetic aquametry: electromagnetic wave interaction with water and moist substances*. Springer Verlag, 2005.

[8] E. Scholz, *Karl Fischer titration: determination of water*. Springer, 1984.

[9] A. Sihvola, "Mixing rules with complex dielectric coefficients," *Subsurface Sensing Technologies and Applications*, vol. 1, no. 4, pp. 393–415, 2000.

[10] "Sartorius Omnimark: Moisture Analyzers. Solids content Analysis." [Online]. Available: http://www.sartorius-omnimark.com/. [Accessed: 06-Oct-2011].

[11] R. Knöchel, W. Taute, and C. Döscher, "Stray field ring resonators and a novel trough guide resonator for precise microwave moisture and density measurements," *Measurement Science and Technology*, vol. 18, p. 1061, 2007.

[12] M. Kent et al., "Intangible but not intractable: the prediction of fish'quality'variables using dielectric spectroscopy," *Measurement Science and Technology*, vol. 18, p. 1029, 2007.

[13] T. Sokoll and A. F. Jacob, "Self-calibration circuits and routines for low-cost measuring systems," *Microwave and Optical Technology Letters*, vol. 50, no. 2, pp. 287-293, Feb. 2008.

[14] T. Sokoll and A. F. Jacob, "A Self-Calibrating Low-Cost Sensor System for Moisture Monitoring of Buildings," in *Microwave Symposium Digest, 2006. IEEE MTT-S International*, 2006, pp. 1584-1587.

[15] T. Sokoll and A. F. Jacob, "In-situ moisture detection system with a vector network analyser," *Measurement Science and Technology*, vol. 18, pp. 1088-1093, Apr. 2007.

[16] "AD8302 | 2.7GHz RF / IF Gain Phase Detector | Log Amps/Detectors | Specialty Amplifiers | Analog Devices." [Online]. Available: http://www.analog.com/en/specialty-amplifiers/log-ampsdetectors/ad8302/products/product.html. [Accessed: 06-Oct-2011].

[17] J. Kerouedan, P. Quéffélec, P. Talbot, C. Quendo, S. De Blasi, and A. L. Bran, "Detection of micro-cracks on metal surfaces using near-field microwave resonators," in *AIP Conference Proceedings*, 2009, vol. 1096, pp. 386-393.

[18] B. Kapilevich, A. Lipsky, and B. Litvak, "Application of the cut-off resonator for microwave monitoring of transformer oil," *Microwave and Optical Technology Letters*, vol. 53, no. 1, pp. 66-68, 2011.

[19] S. R. Wylie, A. Shaw, and A. I. Al-Shamma'a, "RF sensor for multiphase flow measurement through an oil pipeline," *Measurement Science and Technology*, vol. 17, no. 8, pp. 2141-2149, 2006.

[20] A. Penirschke and R. Jakoby, "Microwave sensor for accurate material density measurements of gas/solid Flows in pipelines," in *Proceedings of the 36th European Microwave Conference, EuMC 2006*, 2007, pp. 443-446.

[21] N. Klein, H.-J. Krause, S. Vitusevich, H. Rongen, A. Kurakin, and O. N. Shaforost, "Dual-mode microwave cavity for fast identification of liquids in bottles," in *Microwave Symposium Digest (MTT), 2011 IEEE MTT-S International*, 2011, pp. 1-4.

[22] "Emisens: Home." [Online]. Available: http://www.emisens.com/. [Accessed: 04-Oct-2011].

[23] N. C. Wenger and J. Smetana, "Hydrogen Density Measurements Using an Open-Ended Microwave Cavity," *IEEE Transactions on Instrumentation and Measurement*, vol. 21, no. 2, pp. 105-114, May 1972.

[24] U. S. N. A. and S. Administration, *NASA technical note*. National Aeronautics and Space Administration., 1971.

[25] "Instrumentation | Microwave Density Analyzers." [Online]. Available: http://www.toshiba.com/ind/group_display.jsp?id1=9&id2=90. [Accessed: 06-Oct-2011].

[26] E. F. May, T. J. Edwards, A. G. Mann, and D. K. Manning, "Density, dielectric constant and PVT measurements of a gas condensate fluid," *Journal of Petroleum Science and Engineering*, vol. 41, no. 4, pp. 297-308, Feb. 2004.

[27] "Emerson Process Management - Roxar." [Online]. Available: http://www2.emersonprocess.com/en-US/brands/roxar/Pages/Roxar.aspx. [Accessed: 06-Oct-2011].

[28] M. Dressel, O. Klein, S. Donovan, and G. Grüner, "Microwave cavity perturbation technique: Part III: Applications," *International Journal of Infrared and Millimeter Waves*, vol. 14, no. 12, pp. 2489-2517, 1993.

[29] K. Bechgaard, K. Carneiro, M. Olsen, F. B. Rasmussen, and C. S. Jacobsen, "Zero-Pressure Organic Superconductor: Di-(Tetramethyltetraselenafulvalenium)-Perchlorate [(TMTSF)_{2}ClO_{4}]," *Physical Review Letters*, vol. 46, no. 13, p. 852, Mar. 1981.

[30] W. N. Hardy, D. A. Bonn, D. C. Morgan, R. Liang, and K. Zhang, "Precision measurements of the temperature dependence of in YBa2Cu3O6.95: Strong evidence for nodes in the gap function," *Physical Review Letters*, vol. 70, no. 25, pp. 3999-4002, 1993.

[31] A. Porch, J. R. Powell, M. J. Lancaster, J. A. Edwards, and R. G. Humphreys, "Microwave conductivity of patterned YBa2Cu3O7 thin films," *IEEE Transactions on Applied Superconductivity*, vol. 5, no. 2 pt 2, pp. 1987-1990, 1995.

[32] M. J. Lancaster, *Passive Microwave Device Applications of High-Temperature Superconductors*. Cambridge University Press, 2006.

[33] R. R. Mansour, "Microwave superconductivity," *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 750-759, 2002.

[34] M. Nisenoff, "Microwave superconductivity part 1: History, properties and early applications," in *IEEE MTT-S International Microwave Symposium Digest*, 2011.

[35] M. Nisenoff, "Microwave superconductivity - Part 2: Current and future applications," in *IEEE MTT-S International Microwave Symposium Digest*, 2011.

[36] J. Krupka, "Precise measurements of the complex permittivity of dielectric materials at microwave frequencies," *Materials Chemistry and Physics*, vol. 79, no. 2-3, pp. 195-198, 2003.

[37] J. Krupka, "Frequency domain complex permittivity measurements at microwave frequencies," *Measurement Science and Technology*, vol. 17, p. R55, 2006.

[38] J. Liu, C. Chen, H. Lue, and J. Lue, "Measurement of dielectric constants of metallic nanoparticles by a microwave dielectric resonator," *Measurement Science and Technology*, vol. 13, p. 2032, 2002.

[39] J. Wosik, C. Darne, P. Xie, and J. Krupka, "Resonant and Broadband Microwave Characterization of Single-Walled Carbon Nanotubes," in *Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves and Workshop on Terahertz Technologies, 2007. MSMW'07. The Sixth International Kharkov Symposium on*, vol. 1, pp. 85–85.

[40] J. F. Scott, "Applications of Modern Ferroelectrics," *Science*, vol. 315, no. 5814, pp. 954 -959, Feb. 2007.

[41] R. G. Geyer, P. Kabos, and J. Baker-Jarvis, "Dielectric sleeve resonator techniques for microwave complex permittivity evaluation," *Instrumentation and Measurement, IEEE Transactions on*, vol. 51, no. 2, pp. 383–392, 2002.

[42] P. M. Suherman, T. J. Jackson, and M. J. Lancaster, "Comparison of techniques for microwave characterization of BST thin films," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 2, pp. 397-401, 2007.

[43] M. J. Lancaster, J. Powell, and A. Porch, "Thin-film ferroelectric microwave devices," *Superconductor Science and Technology*, vol. 11, no. 11, pp. 1323-1334, 1998.

[44] B. Hu, W. Liu, C. Gao, X. Zhu, and D. Zheng, "Quantitative microscopy of nonlinear dielectric constant using a scanning evanescent microwave microscopy," *Applied Physics Letters*, vol. 89, no. 4, 2006.

[45] D. I. Mircea and T. W. Clinton, "Near-field microwave probe for local ferromagnetic resonance characterization," *Applied Physics Letters*, vol. 90, p. 142504, 2007.

[46] H. Fricke and S. Morse, "The electric capacity of tumors of the breast," *J Cancer Res*, vol. 16, pp. 310–376, 1926.

[47] Xin Wu and O. M. Ramahi, "Near-field scanning microwave microscopy for detection of subsurface biological anomalies," in *IEEE Antennas and Propagation Society International Symposium, 2004*, 2004, vol. 3, pp. 2444- 2447 Vol.3.

[48] E. C. Fear, P. M. Meaney, and M. A. Stuchly, "Microwaves for breast cancer detection?," *Potentials, IEEE*, vol. 22, no. 1, pp. 12–18, 2003.

[49] S. K. Moore, "Better breast cancer detection," *Spectrum, IEEE*, vol. 38, no. 5, pp. 50–54, 2001.

[50] A. Bate, "Note on the whispering gallery of St Paul's Cathedral, London," *Proceedings of the Physical Society*, vol. 50, p. 293, 1938.

[51] G. F. Strouse, "Sapphire Whispering Gallery Thermometer," *International Journal of Thermophysics*, vol. 28, pp. 1812-1821, Oct. 2007.

[52] A. Ibrahim and D. R. S. Cumming, "Passive single chip wireless microwave pressure sensor," *Sensors and Actuators A: Physical*, vol. 165, no. 2, pp. 200-206, Feb. 2011.

[53] R. B. Piejak, "The hairpin resonator: A plasma density measuring technique revisited," *Journal of Applied Physics*, vol. 95, p. 3785, 2004.

[54] R. L. Stenzel, "Microwave resonator probe for localized density measurements in weakly magnetized plasmas," *Review of Scientific Instruments*, vol. 47, p. 603, 1976.

[55] J. Gregório, L. L. Alves, O. Leroy, P. Leprince, and C. Boisse-Laporte, "Microwave microplasma sources based on microstrip-like transmission lines," *The European Physical Journal D*, vol. 60, no. 3, p. 9, 2010.

[56] J. A. . Broekaert, V. Siemens, and N. H. Bings, "Microstrip microwave induced plasma on a chip for atomic emission spectral analysis," *IEEE Transactions on Plasma Science*, vol. 33, no. 2, pp. 560- 561, Apr. 2005.

[57] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of pseudoparticles," *Physical Review Letters*, vol. 38, no. 25, pp. 1440–1443, 1977.

[58] S. Asztalos et al., "SQUID-Based Microwave Cavity Search for Dark-Matter Axions," *Physical review letters*, vol. 104, no. 4, p. 41301, 2010.

[59] R. Bradley et al., "Microwave cavity searches for dark-matter axions," *Reviews of Modern Physics*, vol. 75, no. 3, p. 777, 2003.

[60] C. K. Jen, "A Method for Measuring the Complex Dielectric Constant of Gases at Microwave Frequencies by Using a Resonant Cavity," *Journal of Applied Physics*, vol. 19, p. 649, 1948.

[61] G. Birnbaum, S. J. Kryder, and H. Lyons, "Microwave Measurements of the Dielectric Properties of Gases," *Journal of Applied Physics*, vol. 22, p. 95, 1951.

[62] M. B. Ewing and D. D. Royal, "A highly stable cylindrical microwave cavity resonator for the measurement of the relative permittivities of gases," *The Journal of Chemical Thermodynamics*, vol. 34, no. 7, pp. 1073-1088, Jul. 2002.

[63] S. H. Autler and C. H. Townes, "Stark Effect in Rapidly Varying Fields," *Physical Review*, vol. 100, no. 2, p. 703, Oct. 1955.

[64] N. R. Walker, "New opportunities and emerging themes of research in microwave spectroscopy," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 365, no. 1861, p. 2813, 2007.

[65] S. Y. TANG, Z. N. XIA, Y. J. FU, and Q. GOU, "Advances and Applications of Microwave Spectroscopy," *Chinese Journal of Analytical Chemistry*, vol. 36, no. 8, pp. 1145–1151, 2008.

[66] T. Balle and W. Flygare, "Fabry–Perot cavity pulsed Fourier transform microwave spectrometer with a pulsed nozzle particle source," *Review of Scientific Instruments*, vol. 52, no. 1, pp. 33–45, 1981.

[67] G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, and B. H. Pate, "The rotational spectrum of epifluorohydrin measured by chirped-pulse Fourier transform microwave spectroscopy," *Journal of Molecular Spectroscopy*, vol. 238, no. 2, pp. 200–212, 2006.

[68] A. Hierlemann, "Integrated chemical microsensor systems in CMOS-technology," in *Solid-State Sensors, Actuators and Microsystems, 2005. Digest of Technical Papers. TRANSDUCERS'05. The 13th International Conference on*, 2005, vol. 2, pp. 1134–1137.

[69] S. Doyle, P. Mauskopf, A. Monfarnadini, L. Swenson, S. Withington, and D. Goldie, "Superconducting Lumped Element Kinetic Inductance resonators for millimeter, sub-millimeter and far infrared detection," in *IRMMW-THz 2010 - 35th International Conference on Infrared, Millimeter, and Terahertz Waves, Conference Guide*, 2010.

[70] A. Masood, A. Porch, and D. Barrow, *Microwave Resonators for Highly Sensitive Compositional Analysis*. Lambert Academic Publishing, 2010.

[71] L. Zhang, Y. Ju, A. Hosoi, and A. Fujimoto, "Microwave atomic force microscopy imaging for nanometer-scale electrical property characterization," *Review of Scientific Instruments*, vol. 81, p. 123708, 2010.

[72] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, "Theory of Superconductivity," *Physical Review*, vol. 108, no. 5, p. 1175, Dec. 1957.

[73] "The Nobel Prize in Physics 1972." [Online]. Available: http://www.nobelprize.org/nobel_prizes/physics/laureates/1972/. [Accessed: 06-Oct-2011].

[74] H. Lewis, "Two-Fluid Model of an 'Energy-Gap' Superconductor," *Physical Review*, vol. 102, pp. 1508-1511, Jun. 1956.

[75] A. Porch, P. Mauskopf, S. Doyle, and C. Dunscombe, "Calculation of the characteristics of coplanar resonators for kinetic inductance detectors," *IEEE Transactions on Applied Superconductivity*, vol. 15, no. 2, pp. 552-555, 2005.

[76] "Physics - Content by Unit." [Online]. Available: http://www.learner.org/courses/physics/unit/text.html?unit=8&secNum=4. [Accessed: 06-Oct-2011].

[77] H. Froehlich, "Theory of the superconducting state. I. The ground state at the absolute zero of temperature," *Phys. Rev.;(United States)*, vol. 79, 1950.

[78] P. Monthoux and D. Pines, "Spin-fluctuation-induced superconductivity in the copper oxides: A strong coupling calculation," *Physical review letters*, vol. 69, no. 6, pp. 961–964, 1992.

[79] A. M. Kadin, "Spatial structure of the Cooper pair," *Journal of superconductivity and novel magnetism*, vol. 20, no. 4, pp. 285–292, 2007.

[80] V. F. Weisskopf and E. O. for N. Research, *The formation of Cooper pairs and the nature of superconducting currents*. Cern, 1979.

[81] D. McDonald, "Novel superconducting thermometer for bolometric applications," *Applied physics letters*, vol. 50, no. 12, pp. 775–777, 1987.

[82] P. K. Day, H. G. LeDuc, B. A. Mazin, A. Vayonakis, and J. Zmuidzinas, "A broadband superconducting detector suitable for use in large arrays," *Nature*, vol. 425, no. 6960, pp. 817-821, 2003.

[83] B. A. Mazin, P. K. Day, K. D. Irwin, C. D. Reintsema, and J. Zmuidzinas, "Digital readouts for large microwave low-temperature detector arrays," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 559, no. 2, pp. 799-801, Apr. 2006.

[84] J. Gao et al., "Strongly quadrature-dependent noise in superconducting microresonators measured at the vacuum-noise limit," *Applied Physics Letters*, vol. 98, no. 23, p. 232508, 2011.

[85] J. Gao et al., "A semiempirical model for two-level system noise in superconducting microresonators," *Applied Physics Letters*, vol. 92, no. 21, p. 212504, 2008.

[86] J. Gao, "The physics of superconducting microwave resonators," 2008. [Online]. Available: http://thesis.library.caltech.edu/2530/. [Accessed: 04-Oct-2011].

[87] I. S. McLean, *Electronic imaging in astronomy: detectors and instrumentation*. Springer, 2008.

[88] G. M. Whitesides, "The origins and the future of microfluidics," *Nature*, vol. 442, pp. 368-373, Jul. 2006.

[89] V. Tesař, *Pressure-driven microfluidics*. Artech House, 2007.

[90] S. C. Terry, "A gas chromatography system fabricated on a silicon wafer using integrated circuit technology," Dec-1975. [Online]. Available: http://adsabs.harvard.edu/abs/1975PhDT.....35T. [Accessed: 15-Sep-2011].

[91] A. J. deMello, "Control and detection of chemical reactions in microfluidic systems," *Nature*, vol. 442, pp. 394-402, Jul. 2006.

[92] R. Fair, "Digital microfluidics: is a true lab-on-a-chip possible?," *Microfluidics and Nanofluidics*, vol. 3, no. 3, pp. 245–281, 2007.

[93] H. Becker, "Hype, hope and hubris: the quest for the killer application in microfluidics," *Lab Chip*, vol. 9, no. 15, pp. 2119-2122, Jun. 2009.

[94] E. Livak-Dahl, I. Sinn, and M. Burns, "Microfluidic Chemical Analysis Systems," *Annual Review of Chemical and Biomolecular Engineering*, vol. 2, pp. 325-353, Jul. 2011.

[95] T. Schwalbe, V. Autze, and G. Wille, "Chemical synthesis in microreactors," *CHIMIA International Journal for Chemistry*, vol. 56, no. 11, pp. 636–646, 2002.

[96] K. Saeed, R. D. Pollard, and I. C. Hunter, "Substrate Integrated Waveguide Cavity Resonators for Complex Permittivity Characterization of Materials," *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 10, pp. 2340-2347, Oct. 2008.

[97] S. K. Pavuluri, R. Lopez-Villarroya, E. McKeever, G. Goussetis, M. P. Y. Desmulliez, and D. Kavanagh, "Integrated microfluidic capillary in a waveguide resonator for chemical and biomedical sensing," *Journal of Physics: Conference Series*, vol. 178, p. 012009, Jul. 2009.

[98] H. Lobato-Morales, A. Corona-Chavez, D. V. B. Murthy, and J. L. Olvera-Cervantes, "Complex permittivity measurements using cavity perturbation technique with substrate integrated waveguide cavities," *Review of Scientific Instruments*, vol. 81, no. 6, pp. 064704-064704-4, Jun. 2010.

[99] E. N. Shaforost, N. Klein, S. A. Vitusevich, A. A. Barannik, and N. T. Cherpak, "High sensitivity microwave characterization of organic molecule solutions of nanoliter volume," *Applied Physics Letters*, vol. 94, no. 11, 2009.

[100] E. N. Shaforost, N. Klein, A. I. Gubin, A. A. Barannik, and A. M. Klushin, "Microwave-millimetre wave WGM resonators for evanescent sensing of nanolitre liquid substances," in *European Microwave Week 2009, EuMW 2009: Science, Progress and Quality at Radiofrequencies, Conference Proceedings - 39th European Microwave Conference, EuMC 2009*, 2009, pp. 45-48.

[101] F. Vollmer and S. Arnold, "Whispering-gallery-mode biosensing: label-free detection down to single molecules," *Nat Meth*, vol. 5, no. 7, pp. 591-596, Jul. 2008.

[102] G. R. Facer, D. A. Notterman, and L. L. Sohn, "Dielectric spectroscopy for bioanalysis: From 40 Hz to 26.5 GHz in a microfabricated wave guide," *Applied Physics Letters*, vol. 78, no. 7, pp. 996-998, 2001.

[103] J. C. Booth, N. D. Orloff, J. Mateu, M. Janezic, M. Rinehart, and J. A. Beall, "Quantitative permittivity measurements of nanoliter Liquid volumes in microfluidic channels to 40 GHz," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 12, pp. 3279-3288, 2010.

[104] K. Grenier et al., "Integrated Broadband Microwave and Microfluidic Sensor Dedicated to Bioengineering," *IEEE Transactions on Microwave Theory and Techniques*, vol. 57, no. 12, pp. 3246-3253, Dec. 2009.

[105] J. M. McKee and B. P. Johnson, "Real-time chemical sensing of aqueous ethanol glucose mixtures," *IEEE Transactions on Instrumentation and Measurement*, vol. 49, no. 1, pp. 114-119, Feb. 2000.

[106] L. Li, "Recent development of micromachined biosensors," *IEEE Sensors Journal*, vol. 11, no. 2, pp. 305-311, 2011.

[107] L. Li and D. Uttamchandani, "A microwave dielectric biosensor based on suspended distributed MEMS transmission lines," *IEEE Sensors Journal*, vol. 9, no. 12, pp. 1825-1830, 2009.

[108] Z. Aboush and A. Porch, "Compact, narrow bandwidth, lumped element bandstop resonators," *IEEE Microwave and Wireless Components Letters*, vol. 15, no. 8, pp. 524-526, 2005.

[109] P. Lidström, J. Tierney, B. Wathey, and J. Westman, "Microwave assisted organic synthesis - a review," *Tetrahedron*, vol. 57, no. 45, pp. 9225-9283, Nov. 2001.

[110] I. R. Baxendale, J. J. Hayward, and S. V. Ley, "Microwave Reactions Under Continuous Flow Conditions," *Combinatorial Chemistry & High Throughput Screening*, vol. 10, no. 10, pp. 802-836, 2007.

[111] A. Günther and K. F. Jensen, "Multiphase microfluidics: from flow characteristics to chemical and materials synthesis," *Lab on a Chip*, vol. 6, p. 1487, 2006.

[112] P. Garstecki, I. Gitlin, W. DiLuzio, G. M. Whitesides, E. Kumacheva, and H. A. Stone, "Formation of monodisperse bubbles in a microfluidic flow-focusing device," *Applied physics letters*, vol. 85, p. 2649, 2004.

[113] J. Kim, A. Babajanyan, A. Hovsepyan, K. Lee, and B. Friedman, "Microwave dielectric resonator biosensor for aqueous glucose solution," *The Review of Scientific Instruments*, vol. 79, no. 8, p. 086107, Aug. 2008.

[114] H. Morgan, T. Sun, D. Holmes, S. Gawad, and N. G. Green, "Single cell dielectric spectroscopy," *Journal of Physics D: Applied Physics*, vol. 40, pp. 61-70, Jan. 2007.

[115] E.H. Synge and E. H. Synge, "A suggested method for extending the microscopic resolution into the ultramicroscopic region," *Phil. Mag.*, vol. 6, p. 356, 1928.

[116] R. F. Soohoo, "A Microwave Magnetic Microscope," *Journal of Applied Physics*, vol. 33, p. 1276, 1962.

[117] Z. Frait, "Ferromagnetic resonance of single crystal hexagonal close packed cobalt," *British Journal of Applied Physics*, vol. 15, pp. 993-997, Aug. 1964.

[118] E. A. Ash and G. Nicholls, "Super-resolution Aperture Scanning Microscope," *Nature*, vol. 237, pp. 510-512, Jun. 1972.

[119] D. W. Pohl, W. Denk, and M. Lanz, "Optical stethoscopy: Image recording with resolution λ/20," *Applied Physics Letters*, vol. 44, p. 651, 1984.

[120] A. Lewis, M. Isaacson, A. Harootunian, and A. Muray, "Development of a 500 \AA spatial resolution light microscope:: I. light is efficiently transmitted through [lambda]/16 diameter apertures," *Ultramicroscopy*, vol. 13, no. 3, pp. 227–231, 1984.

[121] T. Wei, X.-D. Xiang, W. G. Wallace-Freedman, and P. G. Schultz, "Scanning tip microwave near-field microscope," *Applied Physics Letters*, vol. 68, p. 3506, 1996.

[122] V. V. Talanov, A. Scherz, R. L. Moreland, and A. R. Schwartz, "A near-field scanned microwave probe for spatially localized electrical metrology," *Applied Physics Letters*, vol. 88, p. 134106, 2006.

[123] M. Tabib-Azar and Y. Wang, "Design and fabrication of scanning near-field microwave probes compatible with atomic force microscopy to image embedded nanostructures," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 52, no. 3, pp. 971–979, 2004.

[124] R. Inoue, Y. Odate, E. Tanabe, H. Kitano, and A. Maeda, "Data analysis of the extraction of dielectric properties from insulating substrates utilizing the evanescent perturbation method," *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, no. 2, pp. 522- 532, Feb. 2006.

[125] A. P. Gregory, *Private communication*, 2010.

[126] J. Lee, C. J. Long, H. Yang, X.-D. Xiang, and I. Takeuchi, "Atomic resolution imaging at 2.5 GHz using near-field microwave microscopy," *Applied Physics Letters*, vol. 97, p. 183111, 2010.

[127] A. G. Ruiter, K. O. van der Werf, J. A. Veerman, M. F. Garcia-Parajo, W. H. Rensen, and N. F. van Hulst, "Tuning fork shear-force feedback," *Ultramicroscopy*, vol. 71, no. 1-4, pp. 149-157, Mar. 1998.

[128] K. Lee, A. Babajanyan, C. Kim, S. Kim, and B. Friedman, "Glucose aqueous solution sensing by a near-field microwave microprobe," *Sensors and Actuators A: Physical*, vol. 148, no. 1, pp. 28-32, Nov. 2008.

[129] B. Friedman et al., "Sensitive, Label-Free DNA Diagnostics Based on Near-Field Microwave Imaging," *Journal of the American Chemical Society*, vol. 127, no. 27, pp. 9666-9667, Jul. 2005.

[130] C. Wall, "AFM-VNA Technique Enables Compound, Calibrated Electrical and Spatial Measurements at Nanoscale," *Microwave Journal*, Dec-2010.

[131] F. M. Serry, "Scanning Microwave Microscope, Agilent application note, 5989-8818EN."

[132] S. M. Anlage, C. P. Vlahacos, S. Dutta, and F. C. Wellstood, "Scanning microwave microscopy of active superconducting microwavedevices," *IEEE Transactions on Applied Superconductivity*, vol. 7, no. 2, pp. 3686-3689, Jun. 1997.

[133] D. J. Barker, P. M. Suherman, T. J. Jackson, and M. J. Lancaster, "Comparison of scanning evanescent microwave microscopy with co-planar waveguide methods of characterization of Ba0.5Sr 0.5TiO3 thin films," in *IEEE International Symposium on Applications of Ferroelectrics*, 2009.

[134] Y. Lu et al., "Nondestructive Imaging of Dielectric-Constant Profiles and Ferroelectric Domains with a Scanning-Tip Microwave Near-Field Microscope," *Science*, vol. 276, no. 5321, pp. 2004 -2006, Jun. 1997.

[135] S. Yun, S. Na, A. Babajayan, H. Kim, B. Friedman, and K. Lee, "Noncontact characterization of sheet resistance of indium-tin-oxide thin films by using a near-field microwave microprobe," *Thin Solid Films*, vol. 515, no. 4, pp. 1354-1357, 2006.

[136] T. A. Sargsyan, "Near-field scanning microwave microscope application on Cu-phthalocyaninethin-film organic field-effect transistor," *Journal of Contemporary Physics*, vol. 46, no. 3, pp. 119-124, 2011.

[137] S. Wu and J.-J. Yu, "Attofarad capacitance measurement corresponding to single-molecular level structural variations of self-assembled monolayers using scanning microwave microscopy," *Applied Physics Letters*, vol. 97, p. 202902, 2010.

[138] G. Cheng, J. Levy, and G. Medeiros-Ribeiro, "Probing microwave capacitance of self-assembled quantum dots," *Applied Physics Letters*, vol. 95, 032103, 2009, DOI:10.1063/1.3184572

[139] R. A. Kleismit, G. Kozlowski, R. Biggers, I. Maartense, M. K. Kazimierczuk, and D. B. Mast, "Characterization of local dielectric properties of superconductor YBa 2Cu3O7-δ using evanescent microwave microscopy," *IEEE Transactions on Applied Superconductivity*, vol. 15, no. 2, pp. 2915-2918, 2005.

[140] A. Tselev, S. M. Anlage, Z. Ma, and J. Melngailis, "Broadband dielectric microwave microscopy on micron length scales," *Review of Scientific Instruments*, vol. 78, no. 4, 2007.

[141] X.-dong Xiang, C. Gao, F. Duewer, H. T. Yang, and Y. Lu, "Analytical scanning evanescent microwave microscope and control stage," U.S. Patent 755096323-Jun-2009.

# 3 Theory

This chapter discusses the theoretical basis of microwave resonant sensors, and introduces the important concepts that later parts of this thesis rest upon. Section 3.1 is a general introduction, and discusses the advantages of using a resonant system to amplify the response of a material to electromagnetic fields. This is followed by a fresh look at resonator modelling, deriving a comprehensive, multi-stage mathematical model of microwave resonant sensors. The development of the resonator model is divided into three major sections:

- Perturbation analysis (Section 3.2) – how the change in resonant frequency and bandwidth can be attributed to the material parameters of a sample;
- Spectral response estimation (Section 3.3) – how measurements of frequency-domain scattering parameters of a sensor are linked to the resonator's 'true' resonant frequency and bandwidth;
- Resonator dynamic response (Section 3.4) – how the response of a resonator to changing environments and real-time measurement conditions be modelled and understood.

In Section 3.5, there is a review of noise and systematic uncertainty in resonant sensor measurements. Finally, in Section 3.6 is a discussion of the various ways to embody and generalise the multidimensional nonlinear functions encompasses aspects of resonant sensor behaviour and their systems. Such functions are more suitable for an automated or embedded sensor system to learn and simulate than the arbitrary mathematical equations historically used for resonator analysis.

## 3.1 Why use a resonator?

A microwave resonator exhibits a change in its stored energy and power loss due to a variation in a physical electromagnetic property (permittivity, permeability, conductivity) of a specific material in a certain region of a resonant system. These properties fully quantify a material's interaction with electric and magnetic fields. The resonant sensor is carefully designed so that only one of these properties may vary at a given time. Then the change in stored energy and power loss in a resonant system can be selectively attributed to the variation in that material property in that specific region – *i.e.,* the resonator acts as a sensor for the change in the material properties of that region.

Often this physical property varies in response to some external factor like photons, chemical composition, temperature, sample size and position, DC magnetic or electric fields, *etc*., which may be the actual objectives of measurement: the desired signal, or measurand. An indirect approach is therefore taken to measure the objective quantity, first by translating it through a physical electromagnetic property and then through a resonant system.

To do this, the resonator is characterised by measuring the amplitudes of electromagnetic waves passed through it, from which its total stored energy and relative power loss are estimated. Then, using knowledge of its electromagnetic standing wave pattern, the variation in the material property, and finally the measurand itself is inferred. This multi-stage process will be broken down and analysed step-by-step in the following sections of this chapter.

The first stage of this process, linking measured amplitudes of electromagnetic waves as a function of frequency, *i.e.,* the resonator spectral response, to its stored energy and power loss, is deeply connected to changes in the characteristic resonance shape. The key resonance characteristics related to energy and power dissipation are location (*i.e.,* resonant frequency) and size (*i.e.,* bandwidth) (Figure 3.1).

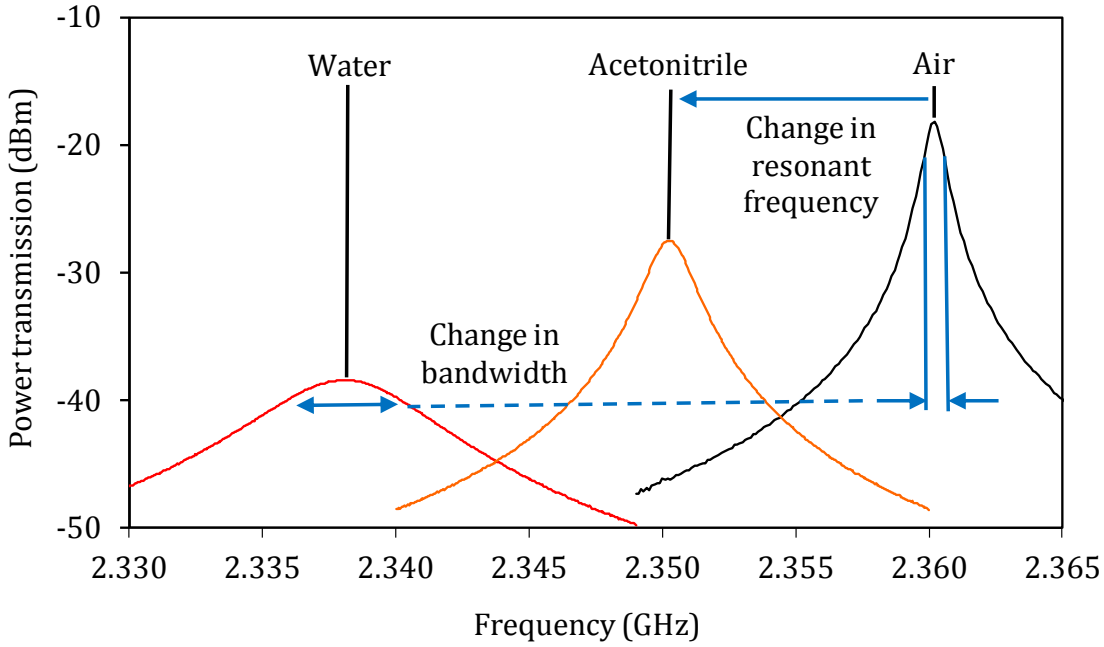


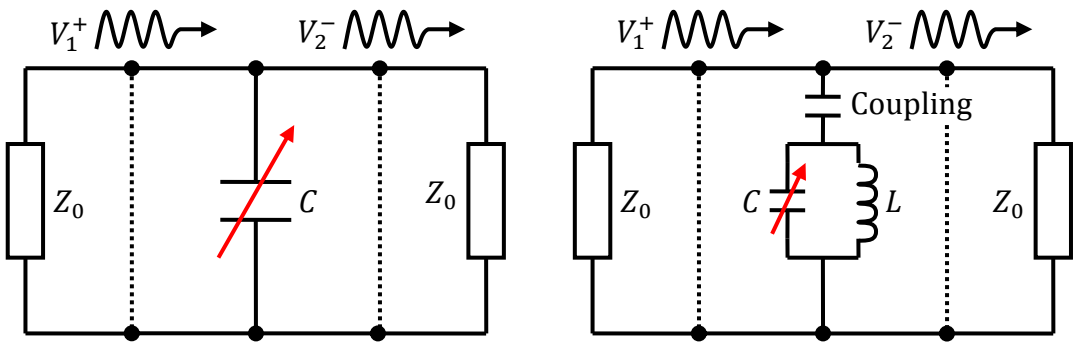

*Figure 3.1: Different substances placed inside a resonator cause both the location of the resonance and its width to be altered – as in this example where different solvents are flowed into a small capillary that passes through a simple metallic rectangular cavity resonator (see Section 6.3 for experimental setup).*

Why take this indirect route? The answer is simple: to achieve a higher sensitivity and selectivity to the objective quantity. In other words, resonators are used to improve the signal to noise ratio (SNR) of the output measurement. The resonator is nothing more than an amplifier – specifically a special type of parametric amplifier. To demonstrate the amplification achieved by a resonant system over a non-resonant one, this discussion starts with a simple MRS measurement example using basic circuit theory.

### 3.1.1 Introductory example of microwave resonant sensing

In this example, two methods of sensing a change in capacitance are compared – a resonant method and a non-resonant method. In a sensor application, this capacitance change would be caused by a change in material properties due to a measurand, for example the presence of a high permittivity, watery human finger –

the basis of the sensors used in touch-sensitive screens. These methods can be represented by the simple circuit diagrams shown below:




Channel select & antialiasing  Simple coupled resonator

*Figure 3.2: Circuit comparison of a resonant (right) vs. non-resonant (left) measurement for determining the change in capacitance, C, by comparing the amplitude and phase of voltage waves measured at the dotted lines.*

To characterise these circuits at microwave frequencies, one can measure the transmission scattering parameter – the ratio of the transmitted 'scattered' voltage wave to the incident voltage wave applied by some generator, assuming both waves are reference to the same system impedance, $Z_0$. This complex-domain scattering parameter is known as $S_{21}$; the subscript denoting the ratio of the outward-travelling voltage wave at port 2, *i.e.,* $V_2^-$, to the inward-travelling voltage wave applied to port 1, *i.e.,* $V_1^+$. See Pozar [1] for definitions and description of scattering parameters in general. In the case of the left circuit containing a single shunt capacitance, $C$, that we wish to measure, $S_{21}$ as a function of the angular frequency, $\omega$, of the voltage waves is given by:

$$S_{21}(\omega) \triangleq \left.\frac{V_2^-}{V_1^+}\right|_{V_2^+=0} = \frac{2}{2 + j\omega C Z_0}. \qquad 3.1$$

If the change in capacitance is very small, the sensitivity of $S_{21}$ to $C$ is

$$\frac{d}{dC} S_{21}(\omega) = -j \cdot \frac{2\omega Z_0}{2 + j\omega C Z_0}. \qquad 3.2$$

This will be the benchmark by which to compare the alternative, resonant method.

The transmission parameter close to resonance of the resonator (right circuit) is

$$S_{21}(\omega) = \frac{K_1 Q_1}{1 + j \cdot 2Q_1 \cdot \frac{\omega - \omega_1}{\omega_1}}. \qquad 3.3$$

This function traces a circle in the complex plane, and its magnitude looks like Figure 3.1. The angular frequency at which this resonator is found to resonate (at the peak in magnitude, where $\Im S_{21}(\omega) = 0$) is denoted by $\omega_1$. The parameter $Q_1$ is the quality factor – this is measure of how sharp a resonance is, and is the ratio of resonant frequency to the 3 dB bandwidth. Both these are 'loaded' parameters. This means after coupling the resonator (with the smaller 'coupling' capacitance at the top in Figure 3.2) to the measurement system impedance. I denote loaded parameters with a subscript '1'.

Generally, loaded parameters will be slightly different from those of the pure, 'unloaded' system, $\omega_0$ and $Q_0$ (*i.e.,* of the isolated parallel LC resonant circuit here). The parameter $K_1$ is related to the strength of the coupling (in this example, this means how large the coupling capacitance is relative to $C$). This parameter will be discussed later on, so for now consider this an arbitrary scaling factor. For the sake of comparison, in this example the resonator consists of a lumped capacitance $C$ and inductance $L$. Then the sensitivity of $S_{21}$ to a small change in the capacitance $C$ on resonance (*i.e.,* $\omega = \omega_1$) would be

$$\frac{d}{dC} S_{21}(\omega = \omega_1) = -j \cdot \frac{K_1 Q_1^2}{C} \quad \left(\omega_1 = \frac{1}{\sqrt{LC}}\right). \qquad 3.4$$

This result ignores the change in $Q_1$ with $C$, which is a second-order effect. The resonator impedance, $Z_R = \sqrt{L/C}$. For the sake of comparing this sensitivity with that of the non-resonant capacitor, if the resonator has $Z_R \approx Z_0$ (it is difficult to create impedances more than an order of magnitude away from 50 Ω at high frequencies anyway), and the comparison is at the same frequency, $\omega = \omega_1 = 1/\sqrt{LC}$, then the single-capacitor sensitivity reduces to:

$$-j \cdot \frac{2\omega Z_0}{2 + j\omega C Z_0} \rightarrow -j \cdot \frac{2\omega_1 Z_R}{2 + j\omega_1 C Z_R} = -\frac{8 + j6}{25\ C}, \qquad 3.5$$

the magnitude of which is $2/5C$. Therefore, at the resonant frequency at least, the magnitude of the resonant sensor's sensitivity is larger by a factor of $2K_1Q_1^2/5$:

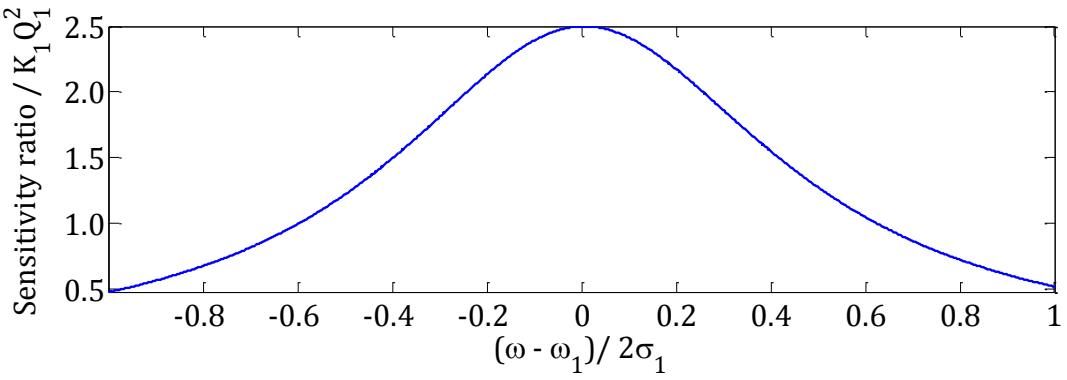



*Figure 3.3: Sensitivity gain of a resonant sensor over a non-resonant sensor as a function of frequencies close to resonance.*

It is apparent that the 'gain' of a resonant sensor is approximately proportional to its quality factor. Just by adding two circuit components, an inductor and a coupling capacitor, a sensitivity gain typically of two or three orders of magnitude is possible – even higher in the case of superconductor resonators. Interestingly, this advantage only appears if the resonator is decoupled. A similar derivation shows that the sensitivity of a parallel LC resonator with no decoupling is 5/4 times that of the single capacitor under the same assumptions – offering no major gain and no advantage. Thus, there is also an inherent assumption here that resonant sensors must be coupled. This occasionally gets 'forgotten' in the literature; in fact the first incarnation of the kinetic inductance detector was not decoupled and hence gave no sensitivity advantage over simply shunting the inductance to ground (although in this case there were other good reasons for using a resonant circuit besides sensitivity) [2].

Coupling reduces the loaded $Q_1$ in a simple way relative to the resonator's inherent unloaded $Q_0$, essentially a fixed parameter determined by the quality of the resonator construction and materials:

$$Q_1 = \frac{Q_0}{1 + |K_1|Q_0}, \quad 0 < |K_1|Q_0 < 1 \tag{3.6}$$

where $|K_1|$ is a measure of the coupling magnitude (this parameter can have an associated phase shift too). By adjusting the coupling, one can (in theory) set $|K_1|$ to any value $|K_1| < Q_1$. The magnitude of $K_1 Q_1$ cannot be greater than one or the circuit would have gain. This would seem to suggest that, in order to achieve maximum sensitivity, the coupling should be made as weak as possible so $|K_1| \ll 1$ and $Q_1 \approx Q_0$. However, whilst this may optimise how narrow the resonance appears, it is not an optimum level of coupling for measurement. Weaker coupling may increase sensitivity up to a point, yet it also reduces the strength of the microwave signal that passes through the resonator; from inspection of Equation 3.3, it is apparent that the relative transmission gain (*i.e.,* insertion loss) on resonance, when $\omega = \omega_1$ is $S_{21}(\omega_1) = K_1 Q_1$. Since any attenuation of a signal degrades its SNR, using weaker coupling also implies that there will be more noise present on the signals transmitted through the resonator, hence also on the resulting estimates of $S_{21}$ and, ultimately, those of the measurand itself.

Optimum sensitivity is thus achieved when the coupling is adjusted so $|K_1|Q_0 = 1$. This is a 'critical coupling' condition when the loaded $Q$ is half that of the unloaded $Q$. Substituting Equation 3.6 into the sensitivity (Equation 3.4) gives $|dS/dC| = |K_1|Q_0^2/(1 + |K_1|Q_0)^2$, which is maximised when $|K_1| = 1/Q_0$. This results in an optimal insertion loss of 6 dB. However, coupling as strong as this may create other issues. For example, the change in loaded bandwidth as measured differs from the change in unloaded bandwidth as coupling is increased; this needs to be corrected for accurate sensing (as discussed later on), which may be a source of error if the correction required is significant. In addition, whatever circuit structures are used for coupling may perturb the internal standing wave field distribution more – possibly making inversion models inaccurate. Furthermore, stronger coupling implies higher sensitivity of the resonance to the external impedance environment – something that may be subject to fluctuations due to temperature, connection/disconnection cycles, *etc.,* as discussed in [3]

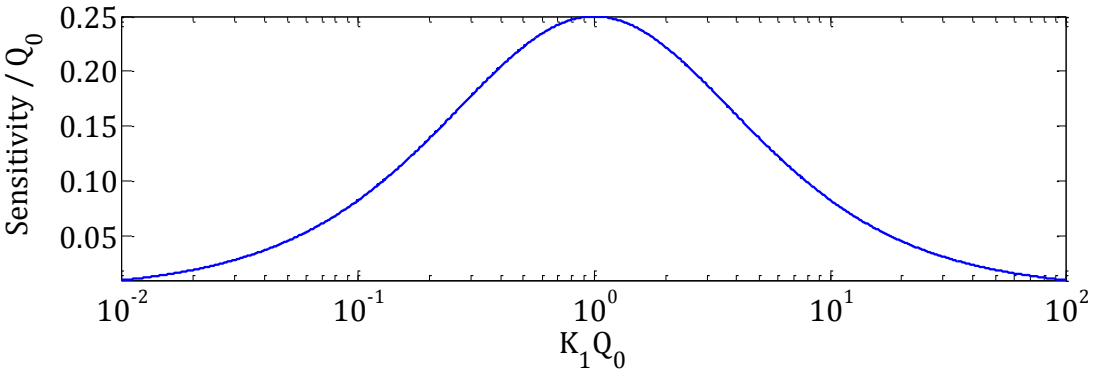



*Figure 3.4: Plot of normalised sensitivity to a change in capacitance (valid also for change in resonant frequency), against coupling factor $K_1Q_0$ (log).*

What this analysis does show, however, is that accepted practice can sometimes be misguided; weak coupling is almost always used 'by default' in the literature, regardless of sensitivity considerations. This analysis has tried to show that resonant sensors can be viewed just like any other type of sensor – and the same considerations apply as to any signal measurement problem. All measurements must balance three central issues: noise, gain, and distortion. This example has shown at a simplistic level how the first two issues relate to MRSs. The complex and tricky issue of distortion – *i.e.,* nonlinearity of sensor response – will also feature many times over the coming discussion, and its minimisation and mitigation is a key problem tackled from various angles throughout this thesis.

### 3.1.2 Introduction to resonant sensor modelling

Like amplifiers, resonant sensors have a nonlinear response to their inputs; they introduce distortion. They also add noise. Also like amplifiers, there is a trade-off between these. Increased gain may improve SNR but at the expense of increased distortion. When talking about sensors, it is more usual to use terms like 'sensitivity' rather than 'gain', but they are really words for the same thing.

A poor but simple approach is to reduce the sensitivity (by introducing a smaller quantity of sample, for example) to a level where one can ignore the nonlinearity altogether, analogous the practice of 'backing off' the input power to an amplifier. This makes processing information in the sensor output easy, but leads to a worse

noise floor and therefore a worse limit of detection (LoD) – the smallest measurand value the system can reliably measure. To an extent this can be improved with higher quality (and higher price) system components, but this results in an undesirable cost-benefit trade-off. An alternative approach is to use signal processing to correct the nonlinearity, and to direct all design efforts towards optimising the gain of the sensor instead. This shifts the onus of the task from good system design to good sensor modelling, as a poor nonlinear model will not fully eliminate the distortion, leading to measurement error.

To this end, I developed a multi-stage modelling approach for resonant sensors. At each stage, signals are defined as functions of time in anticipation of a high-speed readout system and realtime digital signal processing approach. Stage 1 is application-specific, and is the transformation from an objective signal (photon flux, chemical composition, temperature) to one of the material physical parameters. Stage 2 is entirely general, and is the transformation from physical parameters in various regions of the resonator to its defining parameter; its eigenvalue, or complex resonant frequency, $p \triangleq \sigma_0 + j\omega_0$. This parameter is linked to the conventional unloaded resonant frequency and bandwidth of an ideal resonator by the zeros of the conjugate equation, $(s-p)(s-p)^* = 0$, where $s$ is the Laplace transform variable:

$$p = -\frac{\omega_0}{2Q_0}\left(1 \pm j\sqrt{4Q_0^2 - 1}\right) \cong -\frac{\omega_0}{2Q_0} \pm j\omega_0 = \sigma_0 \pm j\omega_0 \quad (Q_0 \gg 1). \tag{3.7}$$

Hence the unloaded bandwidth, $B_0 = \omega_0/Q_0 \cong -2\sigma_0$. The letter $p$ is used rather than the more common $\lambda$ to avoid confusion with wavelength. Note also that $\sigma_0$ should not be confused with conductivity – it is taken from the term for the real part of the Laplace transform variable. Quite often, it is convenient to define $p \triangleq -\sigma_0 + j\omega_0$ instead, as it avoids the negative sign of $\sigma_0$. Since the real part of the eigenvalue is always negative, due to energy conservation in a passive system, the definition being used should always be clear from the context.

The use of eigenvalue, $p$ rather than resonant frequency, bandwidth, *etc.*, is a deliberate choice. This is a parameter that is not directly observable, but is solely

determined by the energy stored and power lost in a resonant mode – a variable that, in turn, can be directly linked to physical material properties, or estimated by numerical finite element method (FEM) simulation to a high degree of accuracy. (as discussed in a later section).

This parameter is not affected by the proximity of other modes, leakage, coupling reactance or other effects – as are the frequency of maximum insertion loss or the point where the phase of a scattering parameter crosses zero. Rather, these points all form part of the general spectral response of the resonator – something that will be discussed in more detail later. This isolates the eigenvalue model from any assumptions about the shape or nature of the measured spectral response, and allows one the flexibility to explicitly model the relationship between spectral measurements and the unknown eigenvalue of an isolated resonator mode. This generality can then accommodate less 'ideal' resonators, for example, when calibration to a reference place near to the resonator is unavailable or impractical, and the electrical length of the connecting cables causes an unknown phase shift in the measuring scattering parameters.

By decoupling eigenvalue from spectral response, the model for stage 2 can be applied universally to any resonator. Stage 3, therefore, is the transformation from these 'hidden' eigenvalues to the frequency-domain scattering parameters.

Scattering parameters (S parameters) are central to linear microwave analysis and are introduced in every textbook on microwave engineering, *e.g.,* [1]. S parameters provide a description of linear systems that transitions smoothly from low frequency circuit theory to optical systems without loss of accuracy, and provide a model of a device that is independent of the impedances presented to its ports. They describe the amplitude and phase ratios of electromagnetic waves travelling outwards (reflected) to those travelling forwards (transmitted) at various spatial reference planes – known as ports. For $N$ ports there are $N^2$ S parameters to describe the complete set of these ratios, which can be arranged in a square matrix.

For the usual microwave situation of guided waves in transmission lines, these quantities also relate to the voltage measured across the transmission line at those reference planes, as the voltage at any point is the superposition of the reflected and transmitted voltage waves. This voltage is sampled by a down-converting mixer and ultimately converted to the digital information in an MRS 'readout' system (see Chapter 4). Thus, S parameters are used as the final output of the model of the microwave resonator, even though they are only approximately measured by the system proposed in this thesis.

An issue arises, however, about how to represent the time-dependency of these scattering parameters – which are a frequency-domain parameter. A simple solution is to use the time-frequency scattering matrix, $\mathbf{S}(\omega, t)$, with each element of the matrix representing one complex wave ratio in two-dimensions; microwave frequency $\omega$ and 'baseband' time $t$. This time-frequency representation assumes that the rate at which the frequency-domain parameter varies in the time domain is slow relative to the period of the frequencies at which it is specified – perfect for modelling modulated signals. This means that the time-stationary assumption of the Fourier representation is not violated for the fast microwave-domain process. Since the period of microwaves is sub-nanosecond, this is almost certainly true.

To summarise so far, a MRS can be represented as a three-stage process. The first stage is the effect of a time-varying measurand, $\theta(t)$, on the material properties (permittivity, permeability and conductivity) of the materials within the resonator. Although there is normally just one measurand, multiple (parasitic) time-varying effects may be represented by the vector, $\boldsymbol{\theta}(t)$. The second stage is the effect of those material properties on the 'energy budget' of the resonator – the total energy loss and storage per mode. This in turn determines the 'hidden' eigenvalues of the resonator, which are functions of the measurand and therefore time: $p(\boldsymbol{\theta}, t)$. Finally, there is the spectral response of the resonator, described using scattering parameters, which is primarily determined by the eigenvalues of the resonator and gives it corresponding peaks or dips with characteristic widths and amplitudes that will change with time. Since these frequency-domain parameters are also time

varying, a time-frequency representation is used for all the scattering parameters: $\mathbf{S}(\omega, t)$.

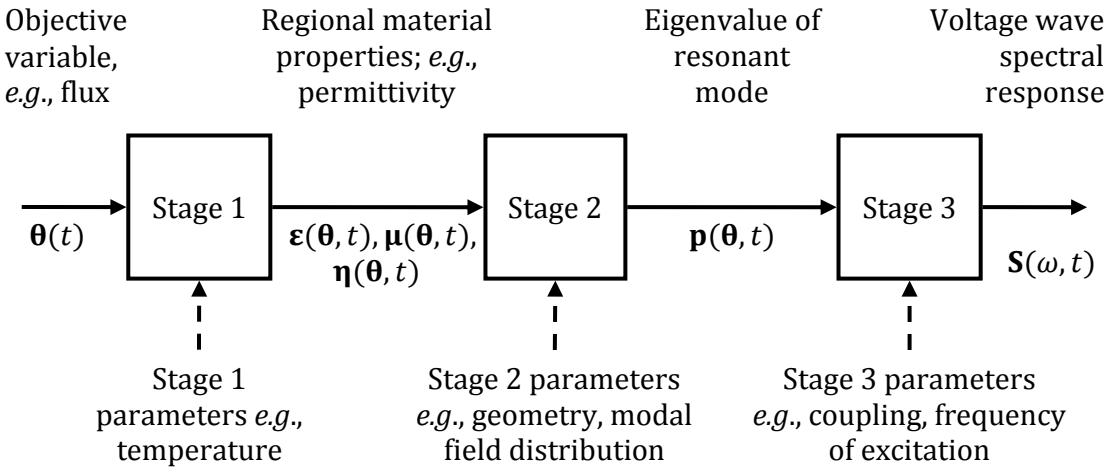



*Figure 3.5: Three stage resonator model showing the main mathematical 'black boxes' involved in the mapping of an objective measurand to the parameter actually measured in any MRS system.*

Stage 1 is specific to each and every application, so if this relationship is known already, the first two stages can be lumped together into one 'resonator eigenvalue response function', which maps directly from the target measurands to the eigenvalues of each resonant mode of interest. Mathematically, this resonator eigenvalue response function (Stage 1+2) can be written:

$$p(\boldsymbol{\theta}, t) = f\big(\boldsymbol{\theta}(t)\big). \tag{3.8}$$

The function $f(.)$ is the response function of the resonator, and the vector $\boldsymbol{\theta}(t)$ contains the time-dependent measurands as discussed above. The vector $\mathbf{p} \in \mathbb{R}^2$can be thought of as the mode's state vector, and over time traces out a trajectory in the two dimensional modal state space.

## 3.2 Resonator perturbation

A resonator can be defined by an energy balance equation, contained within the framework of thermodynamics. Because every system has a Hamiltonian representation, all energy can be seen to come in two 'flavours'; potential energy and kinetic energy [4]. In a system, energy is stored in these forms and continually sloshes between them. This happens because the laws of thermodynamics stipulate that energy must try to distribute itself evenly. Left to its own devices, a system will divide its energy equally, on average, between its potential and kinetic storage mechanisms; the principle known as equipartition. However, the transfer of energy from one form to another is imperfect; it must generate heat and increase its average entropy, fixing the direction of the 'arrow of time' [5]. The condition(s) that guarantee equipartition are known as resonance, and the ratio of the amount of energy stored to the rate at which it is lost from the system due to entropy, governs how quickly the oscillations of a resonator will die away. Normally resonators are driven, however, and thus the energy lost is continually renewed by a supply of new energy from 'outside' the system. Thus, driven resonators can be seen to conserve energy and thus satisfy an energy balance equation without violating the principle of increasing entropy.

Every system, whether microwave, mechanical or chemical, has just three terms in its energy equation – total kinetic energy, total potential energy, and energy dissipation. For example, in kinetic inductance detectors there are three different types of energy storage, magnetic, electric and inertia (of Cooper pairs). Electric energy is a type of potential energy; (PE) and both magnetic and inertial energy are types of kinetic energy (KE). This means that these two mechanisms can be just added up and treated as an 'equivalent' magnetic energy (or inertia, or indeed whatever one wishes), allowing the system to be modelled as a conventional two-mechanism resonator. The standard theory of microwave resonator perturbation comes from Poynting's theorem [6], which is an electromagnetic energy balance equation, but a similar one could be derived for any physical system which has a Hamiltonian that is constant (*i.e.*, conserved) over time. Thus this approach to

resonator modelling is entirely general – for electric energy just read PE, and for magnetic, KE.

The equation most often encountered in traditional perturbation analysis is derived by a number of authors from Poynting's theorem, [1], [7], [8], [9] and in the notation I have adopted looks like this, with $p \triangleq \sigma_0 + j\omega_0$:

$$\frac{p(\theta) - p_0}{p_0} = \frac{\int_{dV} \big(E(\theta) \cdot D^*(0) - E(0) \cdot D^*(\theta)\big) - \big(H(\theta) \cdot B^*(0) - H(0) \cdot B^*(\theta)\big)}{\int_{dV} E(0) \cdot D^*(\theta) + H(0) \cdot B^*(\theta)} \qquad 3.9$$

This says that the relative change in eigenvalue due to a measurand $\theta$ is equal to the relative change in complex power within the resonator volume. Since the equation is in the Fourier domain, a complex number can be used to represent energy storage and power loss of time-harmonic quantities. This time-harmonic *ansatz* is a convenient way of describing energy and power conservation in a single complex equation; all the time-integrals reducing to simple expressions. Since the resonator is driven, both real and imaginary power is conserved (the real power flow out as heat is exactly balanced by real power flow in from the excitation).

Using the electromagnetic constitutive relations $D = \varepsilon E$ and $B = \mu H$, assuming a linear, isotropic (*i.e.,* scalar) material, and using the commutativity of the dot product ($a \cdot b^* = b \cdot a^*$), gives:

$$\frac{p(\theta) - p_0}{p_0} = \frac{-\int_{dV} E(0) \cdot \big(\varepsilon(\theta) - \varepsilon(0)\big)E^*(\theta) - H(0) \cdot \big(\mu(\theta) - \mu(0)\big)H^*(\theta)}{\int_{dV} E(0) \cdot \varepsilon(\theta)E^*(\theta) + H(0) \cdot \mu(\theta)H^*(\theta)} \qquad 3.10$$

The fact that this equation depends on both the unperturbed fields, $H(0)$, and the new changed fields under the influence of a changed sample, $H(\theta)$, is usually ignored in simple perturbation theory, with the assumption that the sample is so tiny as to have no effect of the fields. Whilst patently wrong, this assumption has given acceptable results for many decades. So why go further? The justification for accepting sample-dependent fields and the nonlinearity this causes is made in the context of the type of sensors discussed and the application areas targeted.

Smaller, lighter, and more sensitive sensors mean larger sample volumes and therefore the invalidation of the small perturbation assumption.

This should not be avoided but embraced, as it is not without nonlinearity in the upper measurement ranges that high sensitivity can be achieved in the low measurement ranges; within the limit of the resonator's coupling and minimum loaded Q factor required to resolve the resonance, of course (although automatic coupling control could alleviate this limitation to an extent).

This is a complex domain equation, since not only are the permittivity and permeability complex, but so are the inner products $E(0) \cdot E^*(\theta)$ and $H(0) \cdot H^*(\theta)$, in general. This means that the change in field distribution can cause a change to both the power loss and energy stored – regardless of what is going on inside a perturbing material. The total complex power is given by a volume integral over all space. This is not very practical, and it is logical to break down the integral into a number of discreet regions $i = 1 \ldots n$ of homogeneous material properties:

$$\frac{p(\theta) - p_0}{p_0} = \frac{-\sum_{i=1}^{n} e_i\big(\varepsilon_i(\theta) - \varepsilon_i(0)\big) - h_i\big(\mu_i(\theta) - \mu_i(0)\big)}{\sum_{i=1}^{n} e_i\varepsilon_i(\theta) + h_i\mu_i(\theta)}, \qquad 3.11$$

Where $e_i = \int E(0) \cdot E^*(\theta)\ dV_i$ and $h_i = \int H(0) \cdot H^*(\theta)\ dV_i$ for brevity. Notice the step of abstraction here – from integration to summation. It is a logical extension to say that this equation is a finite element equation; breaking down a problem into a discrete approximation and using summation rather than integration. Thus this equation applies equally well to FEM simulation results when each region is an individual mesh element.

This summation of individual energy packets, possible because of the additive nature of energy itself, allows one to introduce other models of energy storage and loss other than volumetric: the surface of a good conductor, loss due to radiation, a port for coupling, or an isolated electric dipole, for example. Thus, a term might be added to the numerator to account for the change in complex power at the surface of a good conductor:

$$\int H(0)\cdot\big(\eta_k(\theta)-\eta_k(0)\big)H^*(\theta)\,dS_k = h_{S,k}\delta\eta_k(\theta) \qquad 3.12$$

The surface integral is used here, and the impedance $\eta_k \cong (1+j)/\sigma\delta_s$ for a good conductor, resulting in roughly equal perturbation to both energy storage and loss from a change in real surface impedance. In order to avoid added complexity here, the skin depth $\delta_s$ is assumed constant – if the conductivity were the measurand in this sensor then the actual dependence of $\delta_s$ on $\sqrt{\omega}$ would be a source of a small degree of nonlinearity and error if left uncorrected.

This is often a source of nonlinearity (dispersion) in microwave systems and even in FEM eigenvalue simulations, where one approach to correct for $\sqrt{\omega}$ dependence requires making increasingly accurate estimates for the 'true' resonant frequency, putting these estimates back into the equation for skin depth, and iterating. Creating a surface with $\eta = 50\Omega$ can represent a terminated port at the end of a matched transmission line, so this term is also useful for incorporating the effect of loading from a measurement system, such as an additional coupled port.

Dipoles are a mathematical model of an infinitesimal point of charge separation (electrical) or circulating current or spin (magnetic). In resonator analysis, they are frequently used to represent very small samples of spherical or cylindrical geometry. The complex power due to a dielectric dipole is given by:

$$E(0)\cdot p^*(\theta) = V_j\,E(0)\cdot\frac{(\varepsilon_i-\varepsilon_e)}{\varepsilon_e+N(\varepsilon_i-\varepsilon_e)}E^*(\theta) = e_j\tilde{\varepsilon}_j. \qquad 3.13$$

Note $p^*(\theta)$ here is the dipole moment and not the resonator eigenvalue. An equivalent expression exists for magnetic dipoles. Point values for the fields are used rather than surface or volume integrals as an infinitesimal sample is assumed. Here $\varepsilon_i$ and $\varepsilon_e$ are the internal and external complex permittivity, and $N$ is the depolarisation factor; related to the geometry of the sample. $N$ is a scalar quantity for ellipsoidal bodies such as spheres and cylinders. The fraction can be represented by an 'effective' permittivity $\tilde{\varepsilon}$; being the homogeneous material of the same size as the sample which would store the same complex energy in the absence of depolarisation. There is a significantly nonlinear relationship described

by even this very simple depolarisation model – which highlights the importance of accurate nonlinear models and good nonlinear inversion in measurements where depolarisation occurs.

Other expressions for dipole moments exist for small conducting particles and wires, and effective depolarisation factors have been derived analytically or empirically via FEM simulation for a number of other geometries of body besides ellipsoids, *e.g.,* [10], [11], and layered bodies such as a dielectric tube surrounding a dielectric rod [12].

These additional energy contributions can be introduced into the above equation:

$$\frac{p(\theta)-p_0}{p_0}=\frac{-\sum_{i=1}^{n} e_i\Delta\varepsilon_i(\theta)-h_i\Delta\mu_i(\theta)-\sum_{k=1}^{m} h_{S,k}\Delta\eta_k(\theta)-\sum_{j=1}^{p} e_j\Delta\tilde{\varepsilon}_j(\theta)}{\sum_{i=1}^{n} e_i\varepsilon_i(\theta)+h_i\mu_i(\theta)+\sum_{k=1}^{m} h_{S,i}\eta_k(\theta)+\sum_{j=1}^{p} e_j\tilde{\varepsilon}_j(\theta)}. \qquad 3.14$$

In theory, this equation can simulate the resonator eigenvalue response from knowledge only of its particular regions and their field distributions. Unfortunately, pure mathematical knowledge of a resonator's fields, especially how they change if a new sample material is introduced – *i.e.,* the quantity $\big(E(\theta)-E(0)\big)$ – is rarely available. Linear perturbation theory thus makes the assertion that $E(\theta)\cong E(0)$ and $H(\theta)\cong H(0)$ for simplicity. However, this forces the sensor designer to make the volume of sample very small in order to minimise the error it introduces. In turn, this reduces the sensitivity and may increase the physical size of the sensor as well – both undesirable trade-offs.

The origin of the change in fields is three-fold. Firstly, dispersion (frequency dependence of material parameters) may alter the phase constant and attenuation constant of the propagating mode that is resonant. Thus, the propagating mode stays the same, but its energy storage and loss are not exactly what they would be at the original resonant frequency before introduction of a sample.

Secondly, the dominant propagating mode may be joined by another propagating mode – making the resonator multimodal – when a sample is introduced. Energy is partitioned between the modes of propagation via the superposition principle, in a

ratio so as to minimise the free energy. Thus, if a sample makes the original mode energetically unfavourable then an increasing proportion of the energy will be directed into an alternative mode. This happens, for instance, in $TM_{010}$ cylindrical cavity resonators with a lossy dielectric rod down the centre. The 'parasitic' mode in this case is a coaxial TEM mode, where currents (either real or displacement) flow down the rod forming a 'centre conductor' of a coaxial cable and giving rise to a concentrated region of counter-circulating magnetic field and radial electric field close to the rod. This is one of the few cases simple enough to have an analytical description [8].

The third origin of the field change is sample depolarisation/demagnetisation, which creates additional fields due to the accumulation of surface charges and/or circulating currents within an introduced sample. Although accounted for in the ideal dipole case by the above equation, if the dipole fields reach a discontinuity such as a cavity wall or dielectric region other than the sample, then this constitutes a sample-dependent change in field and the energy stored/lost in these regions.

As an example, suppose that the measurand represents the real and imaginary permittivity in one region – the sample. Thus $\varepsilon_i(\theta) - \varepsilon_i(0) = \theta$ for $i = s$ only, where $\theta$ is complex. No other material properties in other regions have any dependence on $\theta$. These equations also apply in situations where the material parameters in just one region are linearly dependent on the measurand. Using the notation derived above, this scenario is described by the perturbation equation:

$$\frac{p(\theta) - p_0}{p_0} = \frac{-e_s(\theta)\theta}{e_s(\theta)\theta + \sum_{i=1, i\neq s}^{n}(e_i(\theta)\varepsilon_i + h_i(\theta)\mu_i)}, \qquad 3.15$$

The 'energy excluding sample' term on the denominator, $\sum_{i=1, i\neq s}^{n}(e_i(\theta)\varepsilon_i + h_i(\theta)\mu_i)$, can be replaced with a single function – $w(\theta)$ – having arbitrary $\theta$ dependence in order to account for sample-dependent fields:

$$\frac{p(\theta)-p_0}{p_0}=\frac{-e_s(\theta)\theta}{e_s(\theta)\theta+w(\theta)}. \tag{3.16}$$

Since it is assumed that this sample-dependent change in fields is quite small, their dependence on some measurand could be expressed by using the first few terms of a Taylor series expansion around the resonator's unperturbed state ($\theta = 0$):

$$w(\theta)=\int_{dV_i}\left(E(0)+\nabla_\theta E\;\theta+\tfrac{1}{2}\theta^T\nabla_\theta^2 E\;\theta+\cdots\right)\cdot E^*(0)=a_i+b_i\theta+\cdots \tag{3.17}$$

A similar expansion could also be used for $h_i(\theta)$, $e_i(\theta)$ and $h_{S,i}(\theta)$ as well. $\nabla_\theta E$ is the Jacobian of the field with respect to $\theta$, and $\nabla_\theta^2 E$ is the Hessian matrix of the field with respect to $\theta$. Note that $a_i$ is real but $b_i$ and any higher order terms are complex. In a later section, graphical examples of these derivative field quantities will be given based on FEM simulations, and the issues in using them to actually model large signal perturbation will be discussed. Note that in this linearisation, $a_i$ is real, but $b_i$ is complex.

Taking just the first order, linear approximation for all these measurand-dependent quantities, *i.e.,* $e_s(\theta)\cong a+b\theta$, and $w(\theta)\cong c+d\theta$, leads to the following equation:

$$\frac{p(\theta)-p_0}{p_0}\cong\frac{-(a+b\theta)\theta}{(a+b\theta)\theta+c+d\theta}=\frac{-(b\theta^2+a\theta)}{b\theta^2+(a+d)\theta+c}. \tag{3.18}$$

The resulting complex quadratic equation has two solutions for $\theta$ given a measurement of $p(\theta)$ and assuming all other quantities are known precisely, making it problematic to use in an automatic sensor without the 'intelligence' to know which solution to accept. However, since this equation is already based on linearised quantities, it makes more sense to expand it as a Maclaurin series about zero (the unperturbed state) as well:

$$p(\theta)\cong p_0\left(1-\frac{a}{c}\theta+\left(\frac{b}{c}-\frac{a^2}{c^2}-\frac{ad}{c^2}\right)\theta^2+\mathcal{O}|\theta^3|\right). \tag{3.19}$$

To second order, we have a real-domain first order term: $a/c$ which is just the standard linear perturbation 'filling factor'; the denominator being equal to four times the energy of the resonator with no sample. The other term is quadratic and is a complex number. Thus, there are still two possible solutions for $\theta$ given $p(\theta)$. However, there is more than one way to linearise an expression. Since the above rational function is complex, a more accurate approach is to equate real and imaginary parts, solve simultaneously, and *then* take the series expansion. This bivariate Maclaurin series expansion up to second order in $\boldsymbol{\theta}$ of Equation (14),:

$$\begin{aligned}\omega(\theta) = \omega_0 - \frac{\omega_0 a}{c}\theta_1 - \frac{\sigma_0 a}{c}\theta_2 + \frac{2\sigma_0(ad_1 + a^2 - b_1 c) - 2\omega_0(d_2 a - b_2 c)}{c^2}\theta_1\theta_2 \\ - \frac{\omega_0(ad_1 + a^2 - b_1 c) - \sigma_0(d_2 a - b_2 c)}{c^2}\theta_1^2 \\ - \frac{\omega_0(ad_1 + a^2 - b_1 c) + \sigma_0(d_2 a - b_2 c)}{c^2}\theta_2^2.\end{aligned} \tag{3.20}$$

And for the bandwidth-related part of the eigenvalue:

$$\begin{aligned}\sigma(\theta) = \sigma_0 - \frac{\sigma_0 a}{c}\theta_1 + \frac{\omega_0 a}{c}\theta_2 - \frac{2\omega_0(ad_1 + a^2 - b_1 c) + 2\sigma_0(d_2 a - b_2 c)}{c^2}\theta_1\theta_2 \\ + \frac{\sigma_0(ad_1 + a^2 - b_1 c) - \omega_0(d_2 a - b_2 c)}{c^2}\theta_1^2 \\ - \frac{\sigma_0(ad_1 + a^2 - b_1 c) - \omega_0(d_2 a - b_2 c)}{c^2}\theta_2^2.\end{aligned} \tag{3.21}$$

In these equations the complex terms $b = b_1 + jb_2$ and $d = d_1 + jd_2$ have been separated into their real and imaginary parts. These expressions are rather unwieldy, and it is debatable whether much is gained from them. To proceed, an even simpler case can be assumed – totally static, sample-independent field distributions, *i.e.*, $b = d = 0$. Equating separately the real and imaginary parts of as before, and leaving out the quadratic terms so that $e_s(\theta) \cong a$, and $w(\theta) \cong c$, gives a pair of equations:

$$\Re\left\{\frac{p(\theta) - p_0}{p_0}\right\} = \Re\left\{\frac{-a\theta}{a\theta + c}\right\}, \qquad \Im\left\{\frac{p(\theta) - p_0}{p_0}\right\} = \Im\left\{\frac{-a\theta}{a\theta + c}\right\}. \tag{3.22}$$

This pair of equations can then be solved simultaneously in terms of $\theta = \theta_1 + j\theta_2$ to obtain a solution that is independent of both $\omega(\theta)$ and $\sigma(\theta)$:

$$\omega(\theta) = \frac{\omega_0(ac\theta_1 + c^2) - ac\theta_2\sigma_0}{a^2\theta_1^2 + 2ac\theta_1 + c^2 + a^2\theta_2^2}, \qquad \sigma(\theta) = \frac{\sigma_0(ac\theta_1 + c^2) + ac\theta_2\omega_0}{a^2\theta_1^2 + 2ac\theta_1 + c^2 + a^2\theta_2^2}. \tag{3.23}$$

An equation in terms of resonator Q can also be derived from these equations:

$$Q(\theta) = \frac{a(2Q_0\theta_1 - \theta_2) + 2cQ_0}{a(4Q_0\theta_1 + 2\theta_2) + 2c}, \tag{3.24}$$

This shows that Q is dependent on a non-trivial combination of $\theta_2$, assumed to be proportional to the loss in the sample, and $\theta_1$; proportional to energy storage.

These are surprisingly complicated results, and demonstrate even in a simple static-field model the inherent interdependence between the real and imaginary parts of the eigenvalue – between resonant frequency and quality factor – properties often treated independently without a second thought. Only with a combination of two of these measurements can a complex material parameter be determined accurately. The inverse of these equations:

$$\theta_1 = -\frac{c}{a} \cdot \frac{\omega(\theta)(\omega(\theta) - \omega_0) + \sigma(\theta)(\sigma(\theta) - \sigma_0)}{\omega(\theta)^2 + \sigma(\theta)^2}, \qquad \theta_2 = \frac{c}{a} \cdot \frac{\sigma(\theta)\omega_0 - \sigma_0\omega(\theta)}{\omega(\theta)^2 + \sigma(\theta)^2}, \tag{3.25}$$

is consistent with the real and imaginary expansion of Equation 3.19 to first order. With the approximation that the Q factor is large (*i.e.*, that $\omega(\theta) \gg \sigma(\theta)$), and the change in resonant frequency is insignificant (*i.e.*, that $\omega(\theta) = \omega_0$), they reduce to the standard linear perturbation equations with $a/c$ being equal to the filling factor.

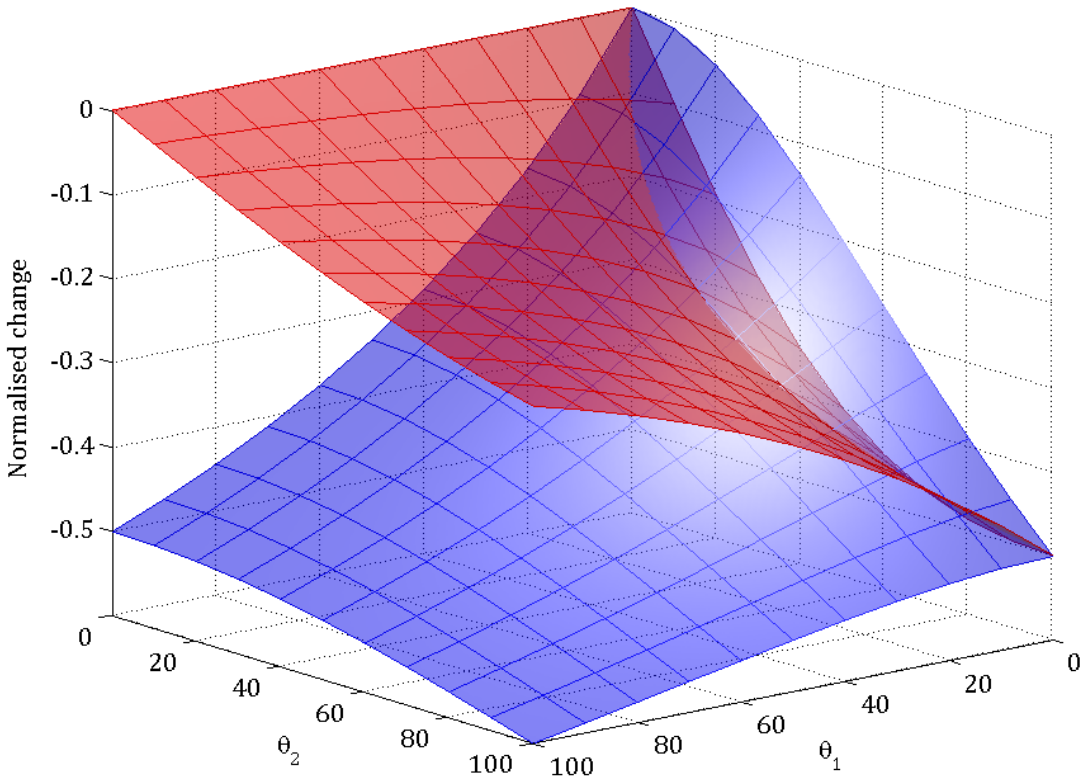



*Figure 3.6: Normalised change* $(\omega(\theta)-\omega_0)/\omega_0$ *(red) and* $(\sigma(\theta)-\sigma_0)/\sigma_0$ *(blue) against real permittivity* $\varepsilon_1-1=\theta_1$ *and imaginary permittivity* $\varepsilon_1=-\theta_2$ *as predicted by Equation 3.23 for a resonator with* $Q_0=1000$ *and filling factor* $a/c=0.01$*. Static field distribution (no depolarisation) is assumed. Although an extreme case, this demonstrates the nonlinear nature of the perturbation response.*

These equations can also be expanded about zero as a truncated Maclaurin polynomial in $\theta$. Retaining just the first order terms in both $\theta_1$ and $\theta_2$ allows a matrix equation to be written, using a 2D vector representation of the complex input variable, $\boldsymbol{\theta}=[\theta_1,\theta_2]^T$ and the complex output variable $\mathbf{p}(\boldsymbol{\theta})=[\sigma(\theta),\omega(\theta)]^T$:

$$\begin{bmatrix}\sigma(\theta)\\ \omega(\theta)\end{bmatrix}\cong\frac{a}{c}\begin{bmatrix}-\sigma_0 & +\omega_0\\ -\omega_0 & -\sigma_0\end{bmatrix}\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}+\begin{bmatrix}\sigma_0\\ \omega_0\end{bmatrix} \qquad 3.26$$

$$or\ \mathbf{p}=\mathbf{K}\,\boldsymbol{\theta}+\mathbf{p}_0$$

This matrix representation is consistent with multiplication by a single complex constant: $p = k\theta + p_0$, since multiplication by $k = a + jb$ is equivalent to multiplication by the matrix $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Clearly, $k = -p_0$, thus $p = -p_0\theta + p_0 = p_0\left(1 - \frac{a}{c}\theta\right)$. This agrees with the first term of the complex Maclaurin series derived above in Equation 3.19 – the matrix representation just serves to highlight the cross coupling that causes correlation between resonant frequency and bandwidth changes in real-life resonator perturbation.

In the usual limit of $\sigma_0 \ll \omega_0$, this reduces to the familiar independent linear perturbation equations – consider for example $\theta_1 = \varepsilon_1$ and $\theta_2 = -\varepsilon_2$; one expects the bandwidth $B(\theta) = -2\sigma(\theta)$ to depend predominantly on $\varepsilon_2$ and resonant frequency $\omega(\theta)$ on $\varepsilon_1$ which would be the case if the diagonal elements of the **K** matrix were zero. The relative magnitude of cross coupling is therefore inversely proportional to the unloaded Q factor of the resonator.

The inverse Equation 3.23 also has a matrix representation. To estimate $\widehat{\boldsymbol{\theta}}$ (in statistics the estimator is donated with a circumflex, or 'hat') from a measured eigenvalue:

$$\begin{bmatrix} \widehat{\theta_1} \\ \widehat{\theta_2} \end{bmatrix} = \frac{c}{a} \cdot \frac{1}{(\omega(\theta)^2 + \sigma(\theta)^2)} \begin{bmatrix} -\sigma_0 & -\omega_0 \\ +\omega_0 & -\sigma_0 \end{bmatrix} \begin{bmatrix} (\sigma(\theta) - \sigma_0) \\ (\omega(\theta) - \omega_0) \end{bmatrix}. \tag{3.27}$$

$$or\ \widehat{\boldsymbol{\theta}} = \mathbf{K}^{-1} \frac{(\mathbf{p} - \mathbf{p}_0)}{|\mathbf{p}|^2}.$$

The above equation, in either matrix or complex form, is recommended for use in any new MRS measurement system where the sample perturbation is small and linear – in the absence of sample-dependent fields in a non-depolarising geometry, it is accurate without further assumptions. Even though ignorance of the cross-coupling terms probably introduces an error of less than 1 % in most situations (*i.e.*, small changes, high Q), as the above equation is still very simple there is no excuse for not using it.

These equations can be thought of as describing 'imperfections' or 'bad design' in resonator responses. However, instead of being considered as such, this non-ideal behaviour should be seen as an inherent – something to be embraced and accepted, rather than avoided and ignored. It is, after all, better to understand, model and account for non-ideal and nonlinear behaviour in a system than to ignore it or try to design it away – leading to inaccuracy or wasted effort, respectively.

### 3.2.1 Sensitivity limits

The result from the previous linear perturbation approximation gives a simple real 'gain' for an MRS sensor, $a/c$. This gain is equal to the filling factor

$$\frac{a}{c} = \frac{\int |E(0)|^2 \, dV_S}{2 \int \varepsilon \, |E(0)|^2 \, dV_R} = \frac{\int |H(0)|^2 \, dV_S}{2 \int \mu \, |H(0)|^2 \, dV_R}. \qquad 3.28$$

Note that the equality of electric and magnetic energy at resonance has been used to simplify the denominator, with the equation in $E$ being for a pure dielectric sample and the equation in $H$ being for a pure magnetic sample. This sensor 'gain' can be seen to be related to the volume ratio of the sample to the resonator, and therefore has a maximum possible value of $1/2$, if all the field were contained in the sample.

As shown earlier, a simple expression for the sensitivity measured by scattering parameters is the tangent-argument of transmission, $\Im\{S_{21}\}/\Re\{S_{21}\}$. Thus the sensitivity to a small change in resonant frequency (rather than capacitance as before) is

$$\frac{d}{d\omega(\theta)} \frac{\Im S_{21}}{\Re S_{21}} \cong \frac{Q_0}{\omega_0(1+g)} \quad (\omega = \omega_0), \qquad 3.29$$

which can be seen to be simply the inverse loaded bandwidth of the coupled resonator.

As derived earlier, the simplest linear model of perturbation says that $\omega(\theta) \cong \omega_0(1 - a\theta_1/c)$, in the absence of any other factors. Therefore, to calculate the overall sensitivity of the resonator, the chain rule can be applied, giving:

$$\frac{d}{d\theta_1}\frac{\Im S_{21}}{\Re S_{21}} \cong -\frac{Q_0}{(1+g)} \cdot \frac{a}{c} = \frac{Q_L a}{c}. \qquad 3.30$$

Thus, in the simplest approximation, the sensitivity of a resonator is the product of the loaded quality factor and the filling factor.

Herein lays the fundamental limit on sensor performance. Suppose a dielectric sample fills the ideal resonator, then $Q_0 \cong 1/\tan\delta$, where $\tan\delta$ is the dielectric loss-tangent of the sample. Assuming coupling is optimised (as discussed earlier), then this situation represents the maximum possible sensitivity. For a high loss material such as water, $\tan\delta \approx 0.1$ at 2.5 GHz. A resonator with a Q of 10 would be almost impossible to couple to and measure. However, decreasing the filling factor to restore Q to a more measureable value reduces sensitivity. Under these simple assumptions, halving the sample volume means doubling the Q but halving the filling factor at the same time, resulting in no net change in sensitivity.

This situation will always approximately apply when the sample is both the objective of measurement and the dominant source of loss – which it often is. For the example of water, therefore, a phase sensitivity of $1/\tan\delta$ radians per 1 part change in sample permittivity represents the best possible performance achievable with a resonant sensor. A 0.1% change in real permittivity results in a maximum of 0.01 radians, or about 0.6°, change in phase. This represents a maximum change of about 0.1 dB in the amplitude of a transmitted signal: a small yet detectable shift.

Similar limits can be shown for other materials, but as the loss in the material decreases so the actual limit of detection achievable is restricted by other factors, such as loss in other parts of the resonator. Although simplistic, what this analysis shows is that detecting small changes in high loss dielectrics should be avoided in MRS, and that even for medium-loss materials there is no significant advantage to be found in extremely high unloaded Q factors.

## 3.3 Resonator spectral response

The resonator spectral response equation, the final stage of the microwave resonator model, has been expressed earlier as $\mathbf{S}(\omega, t) = g(\mathbf{p}, t)$. The nature of this multi-dimensional time-dependent complex equation is the subject of this section. Inverting this relationship, linking measured voltage waves to the resonator eigenvalue, is the most critical step of the microwave resonator readout system. This is especially true in non-VNA based resonator systems, where the absence of reference-plane calibration and precision microwave reflectometers and receivers requires the careful consideration of system imperfections, such as mismatched source and load impedances, to be taken into account.

There are two primary routes to an analytical model of a microwave resonator – a transmission line approximation and a lumped element (LE) approximation. Both these viewpoints are somewhat idealised, and no practical resonator at high frequency can truly be said to belong to either camp entirely. Like all analytical models, these are just approximations of the true subtlety and complexity of reality. However, one of the main applications discussed in the previous chapter, the lumped element kinetic inductance detector (LEKID), is well represented by a LE model, and some effort was expended during this project to derive a representative model of this type of resonant sensor. For this reason, the analytical model of LE resonator spectral response shall be the starting point for this discussion. This is easily generalised, as a lumped component is interchangeable with a distributed reactance at a single frequency, and over small relative frequency deviations (*i.e.,* over a narrow band resonance) the difference is negligible. Indeed, the end results agree with accepted models derived wholly under the assumption of transmission line resonators.

A general model of a resonator must take into account loss in both the magnetic and electric fields; in LE resonators, this means the resistance of the inductor and conductivity of the capacitor dielectric. Notably, the LE circuit is similar to the unit cell of a ‘lossy’ transmission line; thus this circuit (in either parallel or series

format) is an equivalent model for transmission line resonators around a modal frequency. This dual loss model is shown in Figure 3.7.

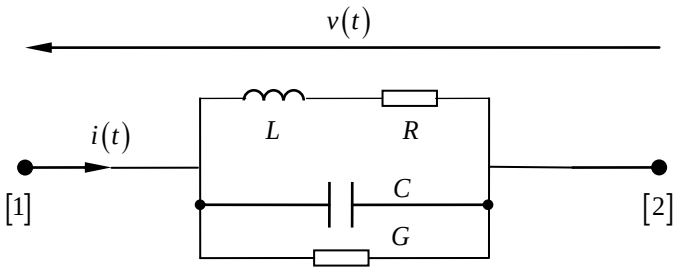



*Figure 3.7: Parallel resonator with dual losses.*

The relationship between the voltage across the circuit and the current through it is given by a differential equation, which can be converted into the Laplace domain, or equivalently by making the *ansatz* that all functions of time $f(t) = z \cdot e^{-st}$ where $z, s \in \mathbb{C}$. The impedance transfer function:

$$Z(s) = \frac{R}{1+RG} \cdot \frac{1 + (L/R)\, s}{1 + \frac{RC+LG}{1+RG} s + \frac{LC}{1+RG} s^2} \qquad 3.31$$

The poles and zeros of this equation yield the conditions under which the resonator acts like an open and short circuit, respectively. These are definitions of electrical resonance if they occur at finite frequency. By comparison to a standard 2nd order form;

$$Z(s) = Z_0 \cdot \frac{1 + a_1 s}{1 + \frac{1}{\omega_0 Q_0} s + \frac{1}{\omega_0^2} s^2}, \qquad 3.32$$

it is possible to write down the complex eigenvalues of the single resonance pole:

$$p = -\frac{\omega_0}{2Q_0}\left(1 \pm j\sqrt{4Q_0^2 - 1}\right) \; where \; \omega_0 = \frac{\sqrt{1+RG}}{\sqrt{LC}} \; and \; Q_0 = \frac{\omega_0 LC}{RC+LG}. \qquad 3.33$$

Note that, since $RG \ll 1, \omega_0 \cong \frac{1}{\sqrt{LC}}$ and that $\frac{1}{Q_0} = \frac{R}{\omega_0 L} + \frac{G}{\omega_0 G} \asymp \frac{1}{Q_c} + \frac{1}{Q_d}$ where $Q_c$ is the 'conductor Q' and $Q_d$ is the 'dielectric Q' – the definition of quality factor implies this rule of additive loss mechanisms. Thus, there are no real surprises with this dual loss analysis when compared to the more common single loss analysis.

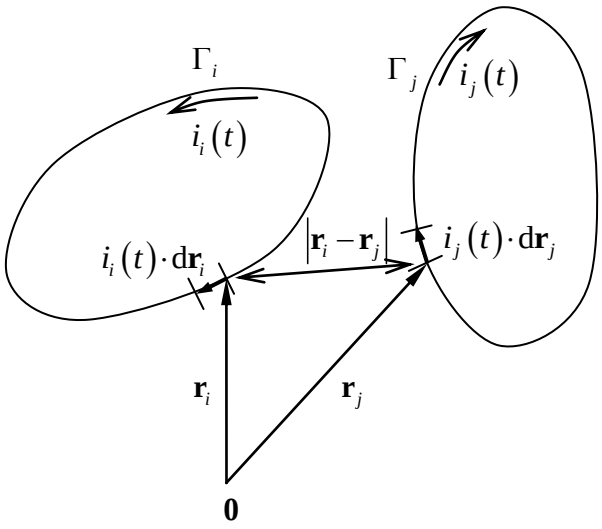
This equation is unrepresentative, as it is virtually impossible to excite such a circuit without loading significantly it by the impedance of the generator and/or the detector used to measure its response. At low frequencies, a voltage source of MΩ impedance could be used, as well as high-impedance oscilloscope, and the circuit model would still be relatively accurate. At RF, 50Ω is king; thus, a form of decoupling (or coupling, depending on your point of view) is needed to avoid this significant loading and associated reduction in resonator selectivity.

This in turn, therefore, requires a new model in order to be representative. Only one type of coupling model will be discussed here. It will be show that the results can be generalised to all coupling methods in general. Again, because of its association with LEKID resonators, the chosen coupling is 'magnetic bandstop coupling'. The term 'bandstop' indicates that the resonator appears as a dip in transmission, as opposed to a peak, and 'magnetic' because the coupling is assumed to be between the resonator's magnetic field and that of the external circuit. It is a somewhat idealised model; a pure LE reactance mutually coupled to the magnetic field of the resonator. However, it is appropriate in the situation where a resonator is in close proximity to an electrically short length of transmission line.

To begin it is necessary to explain the nature of magnetic coupling, which is slightly more complicated than capacitive coupling. Consider the set of *N* electrical circuits, $\{\Gamma_1, \Gamma_2, \dots, \Gamma_N\}$, each carrying a current $i_i$ around the closed contour $\Gamma_i$. The *i*th current element, located at position vector $\mathbf{r}_i$ is $i_i \cdot d\mathbf{r}_i$ (assuming the circuit to be a uniform current filament). The magnetic flux density produced by this element at another point, $\mathbf{r}_j$, is given by the Biot-Savart law, which may be simplified under the uniform current filament assumption:

$$\mathrm{d}\mathbf{B}(\mathbf{r}_j,t)=\frac{\mu_0\mu_r}{4\pi}\frac{\mathbf{J}_i\cdot\mathrm{d}\mathbf{V}_i\times(\mathbf{r}_i-\mathbf{r}_j)}{|\mathbf{r}_i-\mathbf{r}_j|^3}\Rightarrow\mathbf{B}(\mathbf{r}_j,t)=\frac{\mu_0\mu_r}{4\pi}\oint_{\Gamma_i}\frac{i_i(t)\cdot\mathrm{d}\mathbf{r}_i\times(\mathbf{r}_i-\mathbf{r}_j)}{|\mathbf{r}_i-\mathbf{r}_j|^3} \quad 3.34$$

Thus mutual inductance between two circuits, $M_{ij}$, defined as the flux produced by $\Gamma_i$ divided by the current induced by this flux in, $\Gamma_j$, can be derived:




$$\begin{aligned}
M_{ij} &= \frac{\Phi_i(t)}{i_j(t)} \\
&= \frac{1}{i_j(t)}\int_{S_i}\mathbf{B}_j\cdot dS_i \\
&= \frac{1}{i_j(t)}\int_{S_i}\nabla\times\mathbf{A}_j\cdot dS_i, \\
\mathbf{A}_j(\mathbf{r}_i) &= \frac{\mu_0\mu_r}{4\pi}\oint_{\Gamma_j}\frac{i_j(t)\cdot d\mathbf{r}_j}{|\mathbf{r}_i-\mathbf{r}_j|}\Rightarrow \\
M_{ij} &= \frac{\mu_0\mu_r}{4\pi}\oint_{\Gamma_i}\oint_{\Gamma_j}\frac{d\mathbf{r}_i\cdot d\mathbf{r}_j}{|\mathbf{r}_i-\mathbf{r}_j|}.
\end{aligned}$$

*Figure 3.8: Mutual inductance and the derivation of the Von Neumann formula.*

This result is the Von Neumann formula. Note that this equation is general, but for $i=j$, *i.e.,* for self-inductances, the current density and finite cross-section of the conductor needs to be taken into account to avoid the singularity in $|\mathbf{r}_i-\mathbf{r}_j|$. This law applies equally to distributed resonators, although the 'circuit' encompasses a distributed current density. Thus, it is possible to calculate the mutual inductance between, say, a cavity's magnetic field and a coupling loop introduced into it using a similar approach.

Let a system consist of two circuits, $\Gamma_1$ and $\Gamma_2$. If $i_1$ and $i_2$ have the same polarity (as is indicated by the dot notation), the flux in each circuit is given by:

$$\Phi_1(t)=L_1i_1(t)+M_{12}i_2(t),\ \ \Phi_2(t)=L_2i_2(t)+M_{21}i_1(t). \quad 3.35$$

Faraday's law of induction gives the voltage across $L_1$ and $L_2$:

$$v_1(t) = \frac{d\Phi_1(t)}{dt} = L_1\frac{di_1(t)}{dt} + M_{12}\frac{di_2(t)}{dt},\quad v_2(t) = L_2\frac{di_2(t)}{dt} + M_{21}\frac{di_1(t)}{dt}. \qquad 3.36$$

Using the same harmonic *ansatz* used earlier, this may be expressed in terms of the Laplace variable; $V_1(s) = sL_1I_1(s) + sM_{12}I_2(s)$, and $V_2(s) = sL_2I_2(s) + sM_{12}I_1(s)$. Expressing the primary circuit impedance as $V_1(s)/I_1(s)$ and arbitrary secondary circuit impedance such that , gives:

$$Z_1(s) = \frac{V_1(s)}{I_1(s)} = sL_1 - \frac{s^2M_{21}M_{12}}{Z_2(s) + sL_2} = \frac{sL_1Z_2(s) + s^2(L_1L_2 - M_{12}^2)}{Z_2(s) + sL_2}, \qquad 3.37$$

since from inspection of the Neumann formula $M_{ij} = M_{ji}\forall i, j$. In a system of two coupled circuits, $M_{12}M_{21} \leq L_1L_2$ for energy conservation. For finite coupling, $M_{12} = k_m\sqrt{L_1L_2}$, with $k_m = \sqrt{\frac{M_{12}^2}{L_1L_2}}$ a 'magnetic coupling coefficient' satisfying $0 \leq k_m \leq 1$. Another figure used for magnetic coupling is the 'turns ratio' $n = \sqrt{\frac{L_1}{k_m^2L_2}}$. Note that if a sample perturbs neither $L_1$ nor $L_2$, $k_m$ is a constant.

Consider the two extremes of coupling: if $k_m = 0$ the numerator cancels with the denominator leaving just the secondary reactance, $Z_1(s) = sL_1$. Thus, the secondary circuit does not affect the primary at all. Perfect coupling, for $k_m = 1$ gives $Z_1(s) = sL_1Z_2(s)/(Z_2(s) + sL_1)$, which is merely the parallel combination of primary and secondary impedances.

The above relationship can be used to derive the impedance transfer function of a magnetically coupled resonator with dual loss mechanisms as above. Substituting $Z_2(s) = R + 1/(G + sC)$ into the above equation gives:

$$Z(s) = sL_1 \cdot \frac{1 + RG + (RC + L_2(1 - k_m^2)G)\,s + L_2(1 - k_m^2)C\,s^2}{1 + RG + (RC + L_2G)\,s + L_2C\,s^2} \qquad 3.38$$

The denominator of this function can be seen to be the same as that for the uncoupled parallel resonator considered above – its poles are preserved by coupling. The numerator contains a version of the same function, but with $L_2$ reduced by the factor $\left(1 - k_m^2\right)$. Using standard notation:

$$Z(s) = sL_1 \cdot \frac{1 + \frac{1}{\omega_1 Q_1} s + \frac{1}{\omega_1^2} s^2}{1 + \frac{1}{\omega_0 Q_0} s + \frac{1}{\omega_0^2} s^2} \tag{3.39}$$

As well as a conjugate pole pair, the biquadratic fraction also has a conjugate zero pair:

$$z = -\frac{\omega_1}{2Q_1}\left(1 \pm j\sqrt{4Q_1^2 - 1}\right), \qquad \omega_1 = \frac{\omega_0}{\sqrt{1 - k_m^2}}, \qquad Q_1 = \frac{Q_0 - k_m^2 L_2}{\sqrt{1 - k_m^2}} \cong \frac{Q_0}{\sqrt{1 - k_m^2}}. \tag{3.40}$$

The coupled resonator impedance transfer function therefore consists of a conjugate pair of zeros and a conjugate pair of poles, which become closer as the coupling becomes weaker until they cancel completely in the limit of infinitesimal coupling, leaving just the primary circuit reactance. The pole and zero can influence each other; creating a 'skewed' frequency response, but separate into independent resonances as coupling becomes larger. Normally, this is of little practical significance – one resonance becomes dominant and the remaining 'image' resonance has little discernible effect on measured scattering parameters.

The significance of this impedance transfer function to the resonance feature seen in scattering parameters is that both frequency and bandwidth can be seen to change with coupling. This is useful to realise when getting ideal eigenvalue FEM simulations to agree with real-life coupled resonators. It also helps to accurately model the case of a sample-dependent coupling perturbation, whether unintentional or by design (*i.e.,* where the sample perturbs the coupling structure rather than the resonator body in order to achieve a higher sensitivity from the larger 'filling factor' offered by the much smaller coupling structure).

The scattering parameters of this two-port coupled circuit can be found easily by considering the impedance above to be in series with a transmission line:

$$\begin{bmatrix} S_{11}(s) & S_{12}(s) \\ S_{21}(s) & S_{22}(s) \end{bmatrix} = \begin{bmatrix} \frac{Z_1(s)}{2Z_0 + Z_1(s)} & \frac{2Z_0}{2Z_0 + Z_1(s)} \\ \frac{2Z_0}{2Z_0 + Z_1(s)} & \frac{Z_1(s)}{2Z_0 + Z_1(s)} \end{bmatrix} \quad 3.41$$

An analogous expression exits when the impedance is assumed to be in a 'shunt' configuration. The equivalent transformation for a one-port coupled resonator is

$$S_{11}(s) = \frac{Z_1(s) - Z_0}{Z_1(s) + Z_0}. \quad 3.42$$

Taking one of cases as an example, with some simplifying terms the Laplace domain equation for the transmission scattering parameters can be written:

$$S_{21}(s) = \frac{2}{2 + x_c s} \cdot \frac{1 + \frac{1}{\omega_0 Q_0} s + \frac{1}{\omega_0^2} s^2}{1 + \frac{\alpha}{\omega_0 Q_0} s + \frac{\alpha}{\omega_0^2} s^2} = K_0(s) + \frac{K_1(s)}{1 + \frac{\alpha}{\omega_0 Q_0} s + \frac{\alpha}{\omega_0^2} s^2}, \quad 3.43$$

where $\alpha = \frac{2Z_0 + sL_1(1-k_m^2)}{2Z_0 + sL_1}$ is a weakly-$s$-dependent parameter for $L_1 \ll 1$, close to unity for $k_m^2 \ll 1$. The quantities $(s)$ , $K_0(s)$ and $K_1(s)$ change little with frequency close to resonance, thus they can be approximated by complex *constants*:

$$\alpha \cong \frac{2 + jx_c(1 - k_m^2)}{2 + jx_c} = 1 + j\frac{K_1}{K_0},$$

$$K_0 \cong \frac{2}{2 + jx_c(1 - k_m^2)}, \qquad K_1 \cong \frac{-2x_c k_m^2}{(2 + jx_c)\left(2 + jx_c(1 - k_m^2)\right)} \quad 3.44$$

$x_c \cong \omega_0 L_1 / Z_0$ is the normalised coupling reactance at resonance. In the limit of weak coupling $K_0$ is simply the scattering parameter of the coupling reactance alone and $K_1$ goes to zero as the pole and zero cancel out. Because these limits are meaningless, it precludes the assumption of weak coupling coefficient as a route to approximation:

$$\lim_{k_m^2 \to 0} \alpha = 1, \qquad \lim_{k_m^2 \to 0} \mathrm{K}_0 = \frac{2}{2 + jx_c}, \qquad \lim_{k_m^2 \to 0} K_1 = 0. \quad 3.45$$

Having two complex poles at positive and negative frequencies is somewhat redundant and can be represented as a single complex pole at high frequencies; the negative frequency pole is far enough away to have no effect on the positive frequency response close to resonance. Actually the assumption of complex frequency will turn out to be very useful when considering modulated systems – quadrature representation of signals at baseband makes complex frequency physically meaningful. Taking $\alpha$ to be a complex constant ( $s \leftarrow j\omega_0$), this complex pole is:

$$p \cong -\frac{\omega_1}{2Q_1} - j\omega_1 \quad where \quad \omega_1 = \frac{\omega_0}{\Re(\alpha^{-1/2})} \; and \; \frac{1}{Q_1} = \frac{2\Im(\alpha^{-1/2})}{\Re(\alpha^{-1/2})} + \frac{1}{Q_0}\frac{\omega_1}{\omega_0}.$$

$$S_{21}(s) - K_0(s) = \frac{\omega_1^2 K_1(s)}{(s-p)(s-p^*)} \cong \frac{-j\omega_1}{2}\frac{K_1(s)}{(s-p)} = \frac{jQ_1K_1(s)}{1 + \frac{2Q_1}{\omega_1}(s - j\omega_1)} \qquad 3.46$$

It turns out that $\Re(\alpha^{-1/2})$ is slightly more than 1, thus the resonant frequency of the pole is slightly less than that due to the uncoupled resonator. In effect, magnetic coupling reduces, or 'pulls' the effective resonant frequency. This means that FEM eigenvalue simulations will generally overestimate resonant frequencies of magnetically coupled resonators, even in the limit of negligible quantisation and model error.

The coupled, or loaded, Q factor can be seen to be made up of the original unloaded Q, with the correction $\omega_1/\omega_0$ that accounts for frequency pulling, and an external Q factor that arises from the loading of two external 50Ω loads transformed through the coupling. This external Q is therefore defined by:

$$Q_e = \frac{\Re(\alpha^{-1/2})}{2\Im(\alpha^{-1/2})} = \frac{\sqrt{\sqrt{x_c^2+4}\sqrt{x_c^2(1-k_m^2)+4} - x_c^2(1-k_m^2) - 4}}{\sqrt{\sqrt{x_c^2+4}\sqrt{x_c^2(1-k_m^2)+4} + x_c^2(1-k_m^2) + 4}}. \qquad 3.47$$

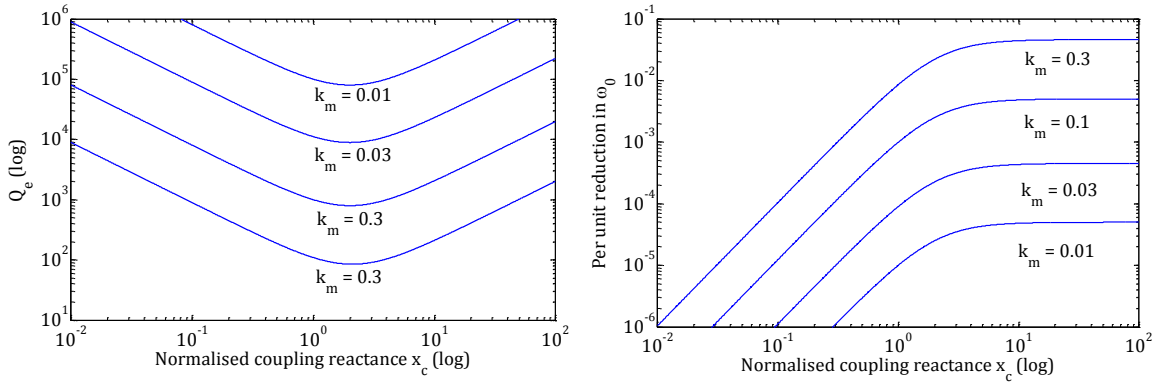



*Figure 3.9: External quality factor of a magnetically coupled bandstop resonator for different mutual coupling strengths and coupling reactance values.*

Therefore, there are two limiting behaviours for external Q in respect to coupling reactance. In the usual case of small coupling reactance it can be seen that $Q_e$ is inversely proportional to both $k_m^2$ and $x_c$. Also, in this limit resonant frequency pulling is proportional to $x_c^2$ and $k_m^2$. However, it is important to remember these simple relationships are only valid in this limit, as shown by the Figure above. This implies that strong coupling cannot be achieved by increasing the size of the coupling structure *ad infinitum.* Maximum potential coupling occurs when the reactance of the coupling structure is equal to the total external load, or 100Ω in this case.

In the small coupling reactance limit, to second order in $x_c$ and $k_m^2$, $\omega_1 \cong \omega_0(1 - x_c^2 k_m^2/8)$ and $Q_e \cong 2/x_c k_m^2$, which agrees with the behaviour seen in this limit above. We can define a coupling coefficient, $g$, which in this limit is therefore

$$g = \frac{Q_0}{Q_e} \cong \frac{Q_0 x_c k_m^2}{2}. \qquad 3.48$$

Also, in the limit of even smaller coupling reactance, *i.e.,* to first order in $x_c$:

$$\alpha \cong 1 - j\frac{x_c k_m^2}{2} = 1 - \frac{j}{Q_e}, \qquad \Re\left(\frac{1}{\sqrt{\alpha}}\right) \cong 1, \qquad \Im\left(\frac{1}{\sqrt{\alpha}}\right) \cong j\frac{x_c k_m^2}{4} \cong \frac{j}{2Q_e},$$

$$K_1 \cong -\frac{x_c k_m^2}{2} = \frac{-1}{Q_e} = \frac{-g}{Q_0}, \qquad K_0 \cong 1 - j\frac{x_c}{2}(1 - k_m^2). \tag{3.49}$$

It is important to note, however, that despite its apparent dependence on $Q_0$, the parameter $K_1$ is still a constant *not* dependent on $Q_0$ in a dynamic sense. This can lead to confusion – hence the choice to leave $Q$ explicitly on the numerator rather than lump it into a new variable parameter. Thus we arrive at the single pole model of the transmission coefficient of a coupled bandstop resonator:

$$S_{21}(s) = K_0 + \frac{K_1 Q_1}{1 + \frac{2Q_1}{\omega_1}(s - j\omega_1)} = K_0 + \frac{K_1 Q_0}{1 + g + \frac{2Q_0}{\omega_1}(s - j\omega_1)} \tag{3.50}$$

Note that this result makes no assumptions about coupling strength – these occur just in the simplified definitions of the parameters. Therefore, it is universally applicable.

The reflection parameters for this resonator are almost exactly the same function, the only difference being $K_0^\Gamma = \frac{jx_c\left(1-k_m^2\right)}{2+jx_c\left(1-k_m^2\right)}$, which is close to zero rather than being close to 1, and $K_1^\Gamma$, which is positive instead of negative:

$$S_{11}(s) = K_0^\Gamma + \frac{K_1^\Gamma Q_0}{1 + g + \frac{2Q_0}{\omega_1}(s - j\omega_1)} \tag{3.51}$$

In fact all simple coupled resonators can be modelled by this function with appropriate choices for the complex parameters $\alpha$, $K_0$, and $K_1$ and thus the derived approximate coupling parameter $g$. This model agrees with the one suggested by Kajfez [13], which is recommended by the UK's National Physical Laboratory (NPL) [14].

In general, accurate values for $x_c$ and $k_m^2$ are not known in advance, and measurements of resonator scattering parameters can only reveal the 'loaded'

eigenvalue $p_L = \sigma_1 + j\omega_1$, This is related to the 'unloaded 'eigenvalue $p_U = \sigma_0 + j\omega_0$ by the following expression:

$$p_L = p_U + \omega_0 a_i + j\omega_0(1 - a_r) \quad \Rightarrow \quad \omega_0 = \frac{\omega_1}{2 - a_r}, \quad \sigma_0 = \frac{\sigma_1(2 - a_r) - \omega_1 a_i}{2 - a_r} \tag{3.52}$$

$$where \;\; a_i = \Im\left(\frac{1}{\sqrt{\alpha}}\right) \;\; and \;\; a_r = \Re\left(\frac{1}{\sqrt{\alpha}}\right).$$

Even if the coefficient $\alpha$ is assumed to be constant, then a change in measured eigenvalue such that $p_L \leftarrow p_L + \Delta p_L$, will not, in general, reflect the true change in unloaded eigenvalue that is predicted by perturbation analysis (or FEM simulation). If uncorrected, this will lead to an error in measurement when attempting to invert the perturbation function. In principle an accurate estimate of $\alpha$ and therefore $a_i$ and $a_r$ could be achieved by determining the constants $K_0$ and $K_1$ and using the identity $\alpha = 1 + j\,K_1/K_0$ (similar identities exist for other types of coupling – see tables below). However, this requires a fully calibrated VNA and the removal of the electrical length of the coupling structures (a post-processing step) in order to give accurate estimates of these parameters (even then, there may be some uncertainty). Fortunately, in the small coupling reactance limit a simple solution is available. To first order in $x_c$, $a_i \cong 1/2Q_e$ and $a_r \cong 1$; thus $\sigma_0 \cong \sigma_1 - \omega_1 a_i$ and $\omega_0 \cong \omega_1$. This ignores frequency pulling due to coupling, although this is an assumption whose accuracy can be verified experimentally by adjusting the coupling mechanically and measuring the perturbation it produces. In this limit:

$$\sigma_0 \cong \sigma_1 - \omega_1 \cdot a_i, \qquad where \qquad a_i \cong \frac{1}{2Q_e} \cong \frac{\left|S_{ij}(\omega_1, 0) - K_0\right|}{2Q_1(0)}. \tag{3.53}$$

The last expression shows how the correction $a_i$ can be estimated from a calibrated scattering parameter measurement of the unperturbed resonator on resonance, $S_{ij}(\omega_1, 0)$, assuming $|K_1| \cong 1/Q_e$. This is valid in principle for any coupled resonator. In terms of loaded and unloaded quality factor, $Q_L$ and $Q_U$, this implies that:

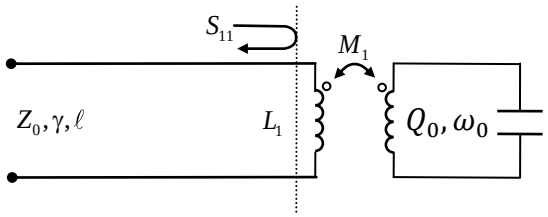
$$Q_U \cong \frac{Q_L}{1 - Q_L|K_1|}. \qquad 3.54$$

Note that this is equivalent to the NPL recommended procedure for obtaining unloaded $Q$ by $Q_U = Q_L/(1-d)$, where $d$ is the diameter of the resonator curve as plotted on a Smith chart (also equivalent to its insertion loss on a linear voltage scale) [14]. It is important to realise that $Q_e$ and therefore $K_1$ are *constant* as long as the coupling does not change and the sample is non-magnetic. Therefore, this correction factor only needs to be evaluated once. The insertion loss measurement $|S_{ij}(\omega_1, 0) - K_0|$ as stated above is valid for the *unperturbed* resonator; it need not be measured with a sample in place. This is valid generally as long as the sample perturbs just one modal field (*i.e.,* magnetic or electric) and the coupling is to the *opposite* field (*i.e.,* magnetic coupling for an E-field sample, capacitive coupling for an H-field sample). As field separation at the sample is normally ensured anyway to avoid ambiguity, this condition is easily met in practice.

Once the constant $a_i$ is known (*e.g.,* through calibrated VNA measurement) it can be applied to all subsequent dynamic acquisitions, even if these are not performed on the VNA. Therefore, the requirement to measure accurate absolute amplitude only applies to an initial 'characterisation' in order to estimate $a_i$, meaning that, thereafter, a simpler hardware without absolute amplitude calibration can be used without introducing error. Clearly, for this assumption to be valid, the coupling strength, coupling reactance and the impedance of the instrument ports must remain static. Lack of precision 50 Ω source and load impedances in alternative readout hardware could be a source of error, therefore. Variations in $Q_e$ due to the sample, *e.g.,* parasitic sample coupling perturbation, depolarisation or imperfect field separation could also invalidate this assumption, as well as temperature variation and mechanical disturbance affecting the external load impedance or coupling reactance. In order to ensure accuracy without requiring absolute amplitude measurements, it is critical that these effects are understood, modelled and corrected for as nonlinearities or mitigated as far as possible.

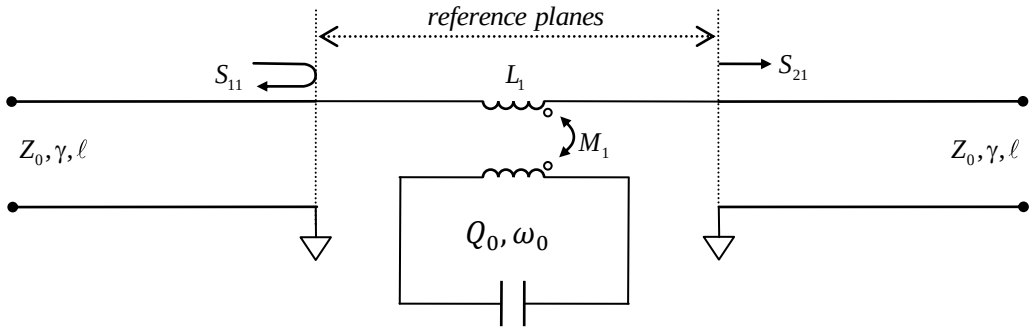
There follows a summary of the various configurations of resonator with appropriate model values for these parameters, showing how reactive coupling affects the complex parameters $\alpha, K_0$, and $K_1$.

### 3.3.1 Magnetic single loop coupling single port reflection:



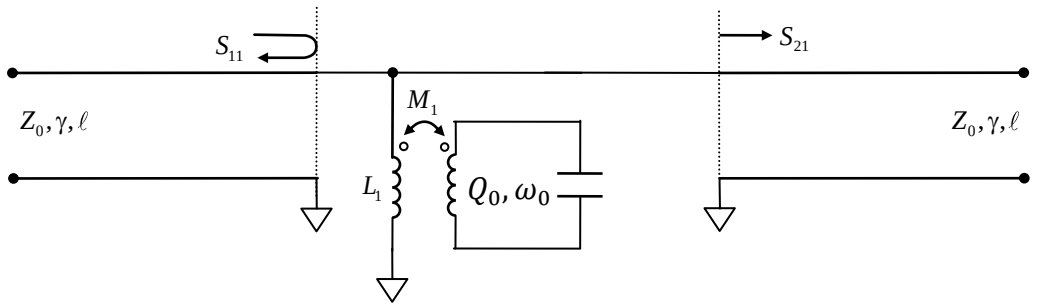

| Parameter | Reflection |
|---|---|
| $\alpha \cong$ | $\frac{1+jx_c(1-k_m^2)}{1+jx_c} = 1+\frac{K_1}{K_0-1} \cong 1-\frac{j}{Q_e} \quad \left(Q_e = \frac{1}{x_c k_m^2}\right)$ |
| $K_0 \cong$ | $\frac{jx_c(1-k_m^2)-1}{jx_c(1-k_m^2)+1} \cong -1+j\left(2x_c-\frac{2}{Q_e}\right)$ |
| $K_1 \cong$ | $\frac{j2x_c k_m^2}{(1+jx_c)\left(1+jx_c(1-k_m^2)\right)} \cong \frac{2}{Q_e}$ |

### 3.3.2 Magnetic single loop coupling bandstop:




| Parameter | Reflection | Transmission |
|---|---|---|
| $\alpha \cong$ | $\frac{2+jx_c(1-k_m^2)}{2+jx_c} = 1 + j\frac{K_1}{K_0-1} \cong 1 - \frac{j}{Q_e}$ | $1 + j\frac{K_1}{K_0} \cong 1 - \frac{j}{Q_e} \quad \left(Q_e = \frac{2}{x_c k_m^2}\right)$ |
| $K_0 \cong$ | $\frac{jx_c(1-k_m^2)}{2+jx_c(1-k_m^2)} \cong j\left(\frac{x_c}{2} - \frac{1}{Q_e}\right)$ | $\frac{2}{2+jx_c(1-k_m^2)} \cong 1 - j\left(\frac{x_c}{2} - \frac{1}{Q_e}\right)$ |
| $K_1 \cong$ | $\frac{2x_c k_m^2}{(2+jx_c)\left(2+jx_c(1-k_m^2)\right)} \cong \frac{1}{Q_e}$ | $\frac{-2x_c k_m^2}{(2+jx_c)(2+jx_c(1-k_m^2))} \cong \frac{-1}{Q_e}$ |

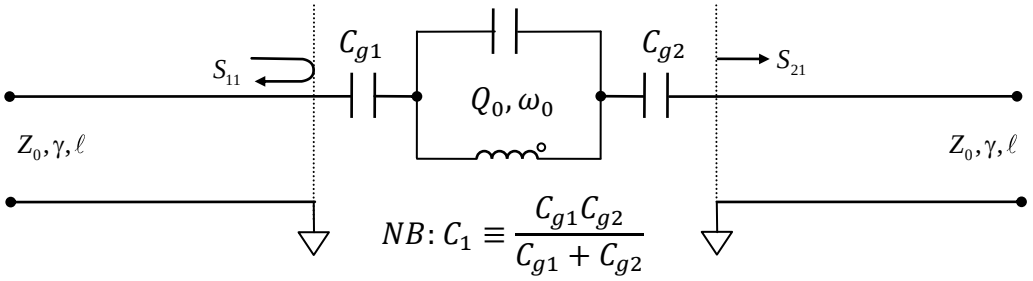
### 3.3.3 Magnetic single loop coupling bandpass:




| Parameter | Reflection | Transmission |
|---|---|---|
| $\alpha \cong$ | $\frac{1+j2x_c(1-k_m^2)}{1+j2x_c} = 1 - j\frac{K_1}{K_0} \cong 1 - \frac{j}{Q_e}$ | $1 + j\frac{K_1}{K_0+1} \cong 1 - \frac{j}{Q_e}\ \left(Q_e = \frac{1}{2x_c k_m^2}\right)$ |
| $K_0 \cong$ | $\frac{-1}{1+j2x_c(1-k_m^2)} \cong -1 + j\left(2x_c - \frac{1}{Q_e}\right)$ | $\frac{-j2x_c(1-k_m^2)}{1+j2x_c(1-k_m^2)} \cong -j\left(2x_c - \frac{1}{Q_e}\right)$ |
| $K_1 \cong$ | $\frac{-2x_c k_m^2}{(1+j2x_c)\left(1+j2x_c(1-k_m^2)\right)} \cong \frac{-1}{Q_e}$ | $\frac{-2x_c k_m^2}{(1+j2x_c)(1+j2x_c(1-k_m^2))} \cong \frac{-1}{Q_e}$ |

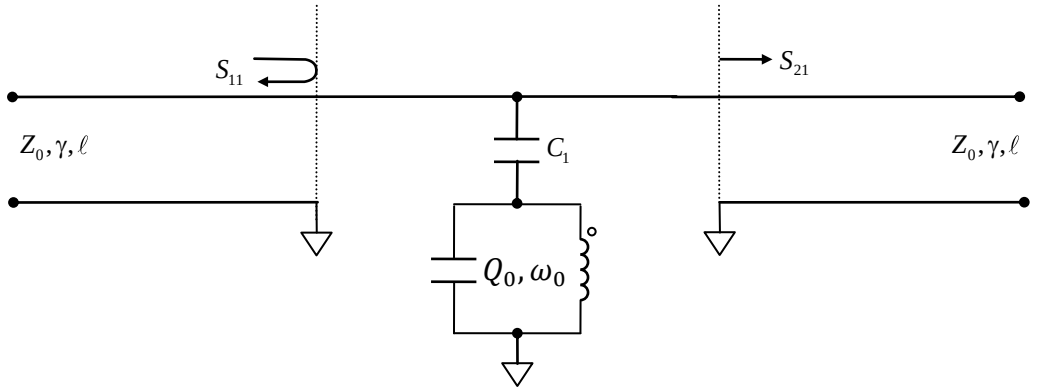
Capacitive coupling is very similar to magnetic coupling; the coupling reactance simply becomes $x_c \cong -1/\omega_0 C_1 Z_0$ and the term $(1 - k_m^2)$ is replaced by $(1 + k_c^2)$, where $k_c = \sqrt{C_1/C_2}$ is an equivalent 'capacitive coupling coefficient'. Note, however, that this term has no upper bound, unlike for magnetic coupling where the term always lies between 0 and 1, as there is no *physical* reason for $C_2$ to be less than the coupling capacitance. Despite this, when the coupling capacitance is small coupling still pulls the resonant frequency down, as $\alpha \cong 1 - j/Q_e$ just as for magnetic coupling.

### 3.3.4 Capacitive coupling bandpass:




| Parameter | Reflection | Transmission |
|---|---|---|
| $\alpha \cong$ | $\frac{2-jx_c(1+k_c^2)}{2-jx_c} = 1 - j\frac{K_1}{K_0} \cong 1 - \frac{j}{Q_e}$ | $1 - j\frac{K_1}{K_0+1} \cong 1 - \frac{j}{Q_e}$ |
| $K_0 \cong$ | $\frac{-2}{2-jx_c(1+k_c^2)} \cong -1 - j\left(\frac{x_c}{2} + \frac{1}{Q_e}\right)$ | $\frac{jx_c(1+k_c^2)}{2-jx_c(1+k_c^2)} \cong j\left(\frac{x_c}{2} + \frac{1}{Q_e}\right)$ |
| $K_1 \cong$ | $\frac{-2x_ck_c^2}{(2-jx_c)\left(2-jx_c(1+k_c^2)\right)} \cong \frac{1}{Q_e}$ | $\frac{2x_ck_c^2}{(2-jx_c)\left(2-jx_c(1+k_c^2)\right)} \cong \frac{-1}{Q_e}$ |

### 3.3.5 Capacitive coupling bandstop:




| Parameter | Reflection | Transmission |
|---|---|---|
| $\alpha \cong$ | $\frac{1-j2x_c(1+k_c^2)}{1-j2x_c} = 1 + j\frac{K_1}{K_0} \cong 1 - \frac{j}{Q_e}$ | $1 + j\frac{K_1}{K_0 - 1} \cong 1 - \frac{j}{Q_e}$ $\left(Q_e = \frac{1}{2x_c k_m^2}\right)$ |
| $K_0 \cong$ | $\frac{-1}{1-j2x_c(1+k_c^2)} \cong -1 - j\left(2x_c + \frac{1}{Q_e}\right)$ | $\frac{-j2x_c(1+k_c^2)}{1-j2x_c(1+k_c^2)} \cong -j\left(2x_c + \frac{1}{Q_e}\right)$ |
| $K_1 \cong$ | $\frac{2x_c k_c^2}{(1-j2x_c)\left(1-j2x_c(1+k_c^2)\right)} \cong \frac{1}{Q_e}$ | $\frac{2x_c k_c^2}{(1-jx_c)\left(1-jx_c(1+k_c^2)\right)} \cong \frac{1}{Q_e}$ |

The case that has not yet been considered is magnetic dual coupling, where there are two magnetic coupling loops and the resonator is placed in series. This is actually the most common configuration. The transmission resonance becomes a peak that is easy to identify and measure. In addition, magnetic coupling loops are easy to manufacture and coupling strength is easy to adjust by rotating the loops. However, a full circuit analysis based on non-ideal transformers (having finite reactance), is too complex to be pursued. However, one can approximate this case with the knowledge that quality factors always combine in a known way. Thus

$$\frac{1}{Q_L} = \frac{1}{Q_0} + \frac{1}{Q_{e,1}} + \frac{1}{Q_{e,2}} \Rightarrow g = g_1 + g_2, \ \ where \ \ g_i = \frac{Q_0}{Q_{e_i}} \cong \frac{Q_0 x_{c,i} k_{m,i}^2}{2}. \qquad 3.55$$

Thus the composite coupling coefficient is simply the summation of the coupling coefficients due to the two (or more) coupling ports. Values for the parameters $\alpha$, $K_0$, and $K_1$ are hard to estimate without a full analysis. However, an approximate derivation using ideal transformers leads to the well-known formulae [15], [16]:

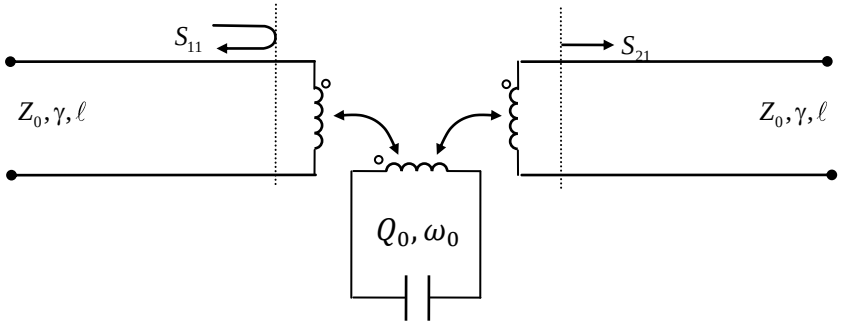
$$S_{11} = \frac{g_1 - g_2 - 1 - j2Q_0x}{g_1 + g_2 + 1 + j2Q_0x} = -1 - \frac{2g_1}{g_1 + g_2 + 1 + j2Q_0x},$$

$$S_{22} = \frac{g_2 - g_1 - 1 - j2Q_0x}{g_1 + g_2 + 1 + j2Q_0x} = -1 - \frac{2g_2}{g_1 + g_2 + 1 + j2Q_0x}, \qquad 3.56$$

$$S_{12} = S_{21} = \frac{2\sqrt{g_1 g_2}}{g_1 + g_2 + 1 + j2Q_0x},$$

where the simplifier $x = (\omega - \omega_0)/\omega_0$. These equations fit into the standard resonator model just derived, thus completing the general resonator model for all possible coupling cases.

### 3.3.6 Magnetic dual bandpass coupling:




| Parameter | Reflection port 1 | Reflection port 2 | Transmission |
|---|---|---|---|
| $\alpha \cong$ | $1 - j/Q_{e,1}$ | $1 - j/Q_{e,2}$ | $1 - j/Q_{e,1} - j/Q_{e,2}$ |
| $K_0 \cong$ | $-1$ | $-1$ | 0 |
| $K_1 \cong$ | $-2/Q_{e,1}$ | $-2/Q_{e,2}$ | $2/\sqrt{Q_{e,1}Q_{e,2}}$ |

### 3.3.7 Modelling spectral response imperfections

To summarise, a good general model of the Laplace domain response of the scattering parameters of a microwave resonator close to one of its resonant modes is:

$$S_{ij}(s) = K_0 + \frac{K_1 Q_1}{1 + \frac{2Q_1}{\omega_1}(s - j\omega_1)} \tag{3.57}$$

This model can also account for a range of non-ideal spectral features. Often it is impractical or impossible to calibrate to a reference plane negligibly far from the coupling structures. Adding a length of transmission line to the coupling structures is equivalent to multiplying the scattering parameters by $e^{-j\omega\tau}$, with $\tau$ equal to the delay of both lines for transmission, and twice that of one line for reflection. This simple phase shift can be seen to substitute $K_0 \leftarrow K_0 e^{-j\omega\tau}$ and $K_1 \leftarrow K_1 e^{-j\omega\tau}$. This is a good reason to consider these parameters to be arbitrary complex constants. If the connecting cables are quite long, linear frequency dependence $K_0 \cong K_0(1 - j(\omega - \omega_0)\tau)$ can be assumed. Short cables can assume a complex constant. Other discontinuities can similarly be approximated by additive and multiplicative complex constants or low-order series expansions in frequency. For the modelling of MKIDs, Gao [17] uses the following analytic model, equivalent under 1st order expansion of $e^{-j2\pi f\tau}$:

$$\hat{S}_{21}(f) = ae^{-j2\pi f\tau}\left(1 + \frac{Q_r/Q_c\, e^{j\phi_0}}{1 + j\frac{2Q_r}{f_r}(f - f_r)}\right), \tag{3.58}$$

where $Q_r \equiv Q_1$ and $Q_c \equiv Q_e$, and $f_r \equiv \omega_1/2\pi$ is the loaded resonant frequency, $a$ is a complex gain-phase error term and $\phi_0$ is an arbitrary phase factor, $\phi_0$, accounts for the electrical length (*i.e.,* reactance) of the 'elbow' coupling structure used in the quarter wavelength superconducting transmission line resonators that Gao used in [17].

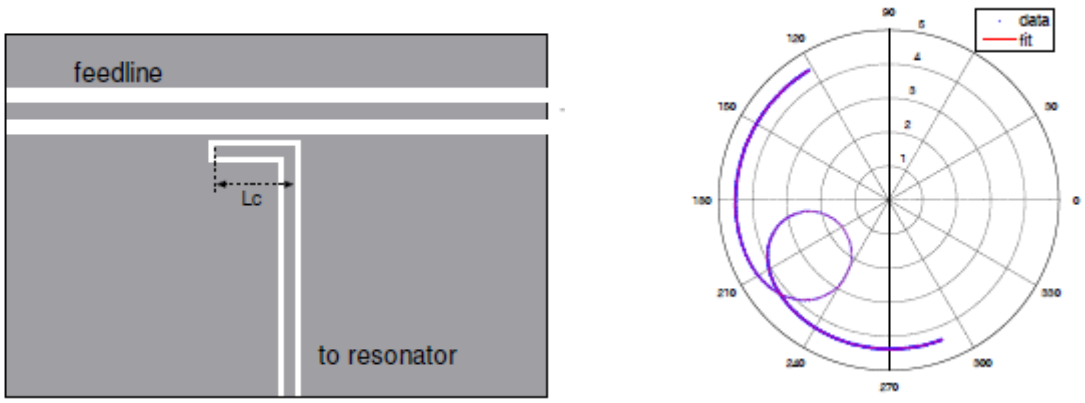



*Figure 3.10: Schematic of the capacitive bandstop 'elbow' coupler of a MKID device (left) and demonstration of the curve fit using the equation above (right). Reproduced from Gao [17].*

Judging by the low fit residual ($\chi^2 = 1.10124$ in this example), this equation is an accurate description of the non-ideal conditions of this MKID resonator. However, to fit this complex 7-parameter equation robustly requires five separate stages – removing the cable delay, circle fitting in the complex plane, rotation and translation, curve fitting to the phase response and then finally deriving all 7 parameters from these fit coefficients, or using these as initial values for the full 7-parameter nonlinear curve fit. Using such a complex procedure is troublesome for an automated MRS system. Apart from the obvious computational burden of this approach, another issue is robustness. Failure to converge to a fit solution can occur due to multiple minima in the error surface of this high-dimensional nonlinear function. During post-processing, this eventuality can easily be discerned by a human operator and corrected by re-fitting or by tweaking the initial parameter estimates. This luxury does not exist in an online sensor processing system.

The power spectral response of this model, taking $s = j\omega$ and $x = (\omega - \omega_0)/\omega_0$, is:

$$\left|S_{ij}(x)\right|^2 = |K_0|^2 + \frac{4Q_0\left(2Q_0\Im(K_1K_0^*)x + Q_0|K_1|^2 + 2\Re(K_1K_0^*)(1+g)\right)}{(1+g)^2 + 4Q_0^2x^2}. \qquad 3.59$$

This is not a Lorentzian function – the term $\Im(K_1 K_0^*) \cong -8x_c g/Q_0$ causes a frequency dependence on the numerator of the function, manifest as a 'skew' in the power transmission. This will prevent an accurate curve fit potentially leading to bias-causing errors. Also, the maximum power transfer is no longer at $x = 0$; causing an error in the measurement of the resonant frequency, although this isn't necessarily an issue if the error is static and differential techniques are employed. To compensate for this error properly requires estimation of the complex parameters $K_0$ and $K_1$, which in turn implies the ability for vector measurement – the measurement of both magnitude and phase response. However, fitting to an empirical 'skewed Lorentzian' function shown in normalised form below, which assumes a linear dispersion model for $K_0(x)$, was shown by Gao to agree with the full complex domain 7-parameter model to within $10^{-7}$ and 0.1% for resonant frequency and quality factor, respectively:

$$\left|S_{ij}(f')\right|^2 = a_1 + a_2 f' + \frac{a_3 + a_4 f'}{1 + a_5 f' + a_6 f'^2} \qquad 3.60$$

In this form, (slightly different to that used by Gao, but with the same number of degrees of freedom), $f'$ should be normalised so that $0 < f' \leq 1$. If this is ensured then the curve fit using this function is quite robust, has reliable convergence, and is computationally efficient – therefore making it suitable for implementation in an online system. This curve fit is general enough to be applied to both transmission and absorption resonances without any further specialisation, although the initial starting points should be chosen appropriately in each case to improve the reliability of convergence. Curve fitting methods are considered in more detail in the Software Development chapter.

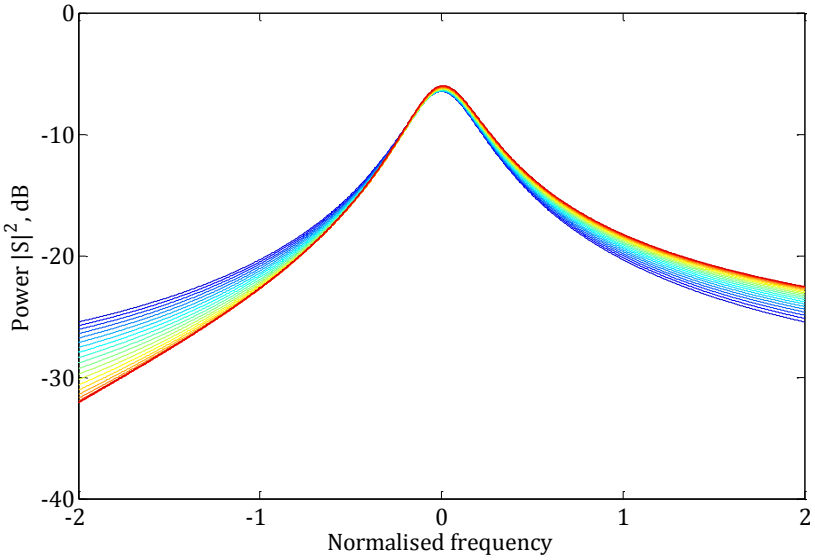



*Figure 3.11: Effect of varying the phase offset $\theta$ between $K_0$ and $K_1$ from $0 \rightarrow 90°$, corresponding to colours blue → red. $K_0 = 0.025e^{j\theta}$, $K_1 = -g/Q_0$.*

Due to inherent properties of coupled resonators, vector measurement is important to ensure accurate results. Although scalar methods – detecting only the power transfer function – seem appealing due to their simplicity, any method aspiring to the accuracy of VNA readout methods should, therefore, ideally allow for the full vector response to be estimated. As will be shown in Chapter 4, modern vector modulation and demodulation techniques facilitate this without greatly increasing system complexity. With complex frequency response data the correction of the error-causing complex parameters $K_0$ and $K_1$ can be done with curve fitting directly to the full complex domain equation, or by correcting for the vector offset and rotation as measured in the complex plane before performing a standard Lorentzian fit to the corrected data.

Another source of error in resonator spectral response measurements that is particularly relevant to non-VNA readout systems is the lack of error-corrected scattering parameters. Full 2-port error correction, which compensates for coupler mismatch, finite directivity and the gain and phase response of the microwave system up to the calibrated reference planes, requires the measurement of all four scattering parameters. Thus, it also requires the use of two receivers, precision reflectometers, switches and other microwave components that are inherently

bulky and expensive. A reduced cost and complexity measurement system cannot afford the luxury of error correction.

For the most part, this is not as serious an issue as it might first seem. One of the main reasons for this is that the resonator, by design, is the highest Q component in the signal chain. What this implies is that the frequency dependence of the connecting cables, *etc.,* can be either taken to be constant, or corrected for by a simple linear dispersion model. This 'background' then has little effect on a proper differential measurement if a simple 'thru' calibration is available. A more serious source of error is uncalibrated source and load mismatches. If the receiver terminates in a mixer, as is usual, then the voltage 'sampled' is the superposition of the forward and reflected voltage waves at the diode. If this termination is non-ideal, then the reflected wave may be re-reflected at the resonator and cause an error. From signal graph theory, the estimated transmission parameter is:

$$\hat{S}_{21} = \frac{b_2}{a_1} = \frac{S_{21}}{1 - S_{22}\Gamma_l} \cong S_{21}(1 + S_{22}\Gamma_l), \tag{3.61}$$

where the last approximation assumes a small mismatch; $\Gamma_l$ is the reflection coefficient of the receiver. Although additive, this error now contains a copy of the resonator-dependent reflection parameter. In the case of an ideal, symmetrically coupled bandpass resonator $S_{22} \cong -S_{21} - 1$. Thus in this case:

$$\hat{S}_{21} \cong S_{21}(1 - \Gamma_l) - \Gamma_l S_{21}^2, \tag{3.62}$$

which can be seen to cause a quadratic distortion to the original transmission. Since generally $S_{22}$ is large when $S_{21}$ is small, and vice-versa, this error will be felt strongest at the edges of resonance, leading to an overestimate of the resonator bandwidth. Again, this may not lead to an error in a fully differential measurement system, but will lead to difficulty in curve fitting based on the assumption of an ideal scattering parameter model. One simple way around this issue, which only requires a modest increase in circuit complexity, is to use a circulator-based isolator on the receiver to terminate the unwanted reflected signal. Since

circulators, especially when narrow band, are relatively compact and moderately inexpensive, this provides a compromise solution.

A more complete error model for transmission measurements [18] which also takes into account source mismatch, crosstalk and gain error (response dispersion) is given by the following equation and signal graph model (Figure 3.12):

$$\hat{S}_{21} = C + \frac{S_{21}T}{1 - S_{11}\Gamma_S - S_{22}\Gamma_L - \Gamma_S\Gamma_L S_{21}S_{12}} \cong C + T(1 - \Gamma_S - \Gamma_L)S_{21} - T(\Gamma_S + \Gamma_L)S_{21}^2, \quad 3.63$$

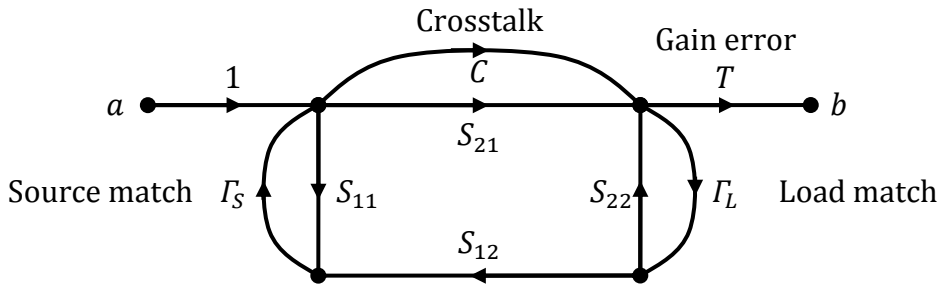



*Figure 3.12: Signal flow graph indicating sources of error in transmission parameter measurement.*

This again leads to a quadratic error when approximated to first order in both source and load mismatch. Note also that, in general, all these error terms are frequency dependent – assuming a constant value for these is based on the narrow band approximation that all these errors have much lower Q factors than the resonator. Whilst clearly the case at microwave, error terms introduced at baseband may well exhibit 'ripple' which cannot be taken as constant across the measurement bandwidth. These issues will be addressed practically in Chapter 4.

The previous discussion has given up expressions describing the static behaviour of microwave resonators. In the next section, the dynamic behaviour will be investigated.

## 3.4 Resonator dynamic response

In a dynamic system, the effect of time-varying parameters must be considered. In the time domain, the resonator is excited with a time-varying electromagnetic signal and the change in that signal in response to a sample-dependent time-varying resonant frequency and bandwidth is measured. The most accurate description of such a system is a time-domain differential equation. This section investigates the behaviour of a resonant sensor subject to rapid variations in its characteristics using differential equation analysis methods.

Starting from the second order system representation derived earlier, assuming $K_0$ and $K_1$ are arbitrary constants, and using the loaded values $\omega_1$ and $Q_1$;

$$S_{ij}(s) = \frac{B_i(s)}{A_j(s)} = K_0 + \frac{K_1}{1 + \frac{1}{\omega_1 Q_1}s + \frac{1}{\omega_1^2}s^2} \tag{3.64}$$

In order to investigate time-dependent behaviour, the first step is to convert this equation to a state space representation, by defining $\mathbf{x}(t) = [x_1, x_2]^{\mathrm{T}}$ as the arbitrary state vector:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \begin{bmatrix} -1/\omega_1 Q_1 & -1/\omega_1^2 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} a_j(t) \\ b_i(t) &= [K_1 \quad 0]\, \mathbf{x}(t) + K_0 a_j(t) \end{aligned} \tag{3.65}$$

With no time-dependence, *i.e.*, if the resonator were a linear time-invariant (LTI) system, this set of equations would be all that is needed to describe its behaviour. This would yield no more useful information than the existing spectral response model derived earlier, however. To introduce parametric variation, it is useful to first rephrase the above state vector equation as second order ordinary differential equation (ODE), by re-substituting the 'dummy variables' $x_2 = x(t)$ and $x_1 = \dot{x}(t)$:

$$\frac{\ddot{x}(t)}{\omega_1^2} + \frac{\dot{x}(t)}{\omega_1 Q_1} + x(t) = a_j(t) \tag{3.66}$$

To tie this together with the perturbation theory described earlier in terms of a changing resonator eigenvalue, consider $\frac{\omega_1}{2Q_1} \leftarrow \sigma_1(t)$ and $\omega_1 \leftarrow \omega_1(t)$. This results in a differential equation with arbitrary time-dependence of its parameters. Furthermore, consider the simple case of the incident voltage wave being a harmonic excitation, $a_j(t) = e^{-j\omega_d t}$:

$$\ddot{x}(t) + 2\sigma_1(t)\dot{x}(t) + \omega_1^2(t)x(t) = \omega_1^2(t) \cdot e^{-j\omega_d t} \qquad 3.67$$

Unfortunately, this deceptively simple linear differential equation hides an extraordinary level of complexity. For example, considering just one form of arbitrary parameter time-dependence gives the Mathieu equation, where a parameter is periodically varying: $\ddot{x}(t) + (a - 2q\cos(2t))x(t) = 0$. The solutions to this equation are in terms of irreducible Mathieu functions [19] and encompass a range of different behaviour, with only certain values for the Mathieu coefficients resulting in periodic output (*i.e.,* chaotic, non-repeating behaviour is the norm). Another related behaviour is observed when the parameters vary at *twice* the resonant frequency, causing amplification of the excitation wave – a phenomenon known as parametric resonance. This is often likened (incorrectly) to the method of self-excitation of a child on a swing [20], the child's legs causing the centre of gravity to be driven at twice the frequency of oscillation thus causing it to swing higher. However, in the case of a small perturbation to the parameters, the behaviour is only weakly different from the solutions for simple harmonic motion.

To help understand the behaviour of this parametrically *perturbed* oscillator, it can be simulated numerically. Consider a sudden change in both parameters at time $t = \tau$, so that $\sigma(t) \leftarrow \sigma + \delta_\sigma u(t - \tau)$ and $\omega(t) \leftarrow \omega + \delta_\omega u(t - \tau)$, where $u(t - \tau)$ is the delayed Heaviside function. The differential equation can be written:

$$\ddot{x}(t) = [\omega + \delta_\omega u(t - \tau)]^2 \left(e^{-j\omega_d t} - x\,(t)\right) - 2[\sigma + \delta_\sigma u(t - \tau)]\dot{x}(t) \qquad 3.68$$

This can be converted into system of first order ODEs by re-assigning dummy variables, which is then amenable to numerical solvers, such as Runge-Kutta methods [21]. Numerical simulation of this system of differential equations is

implemented quite simply in Matlab® (MathWorks Inc., MA, USA). An example is shown below for a step-change of ~ 1% in both $\sigma$ and $\omega$ with the excitation at the unperturbed resonant frequency. After 1000 cycles (long enough for a steady state of oscillation to be reached, the step-change occurs. The results confirm the intuitive results that the resonator takes some time to settle into oscillation at its new resonant frequency. For small changes the resonator acts like an over-damped linear system, and for large changes 'overshoot' or ringing is seen. As shown below, these transients decay exponentially with decay constant $\sim 1/\sigma_1 = 2Q_1/\omega_1$ .

Because the Q is relatively high, the response occurs on a time scale much longer than the oscillation frequency. It is therefore logical to study what happens to its *envelope* with respect to time rather than the voltage wave itself. Indeed, this envelope response is how such a transient will be seen at the baseband level; the response after being mixed with a quadrature carrier equal to the excitation frequency and then low-pass filtered – *i.e.,* a homodyne IQ detection system as used in kinetic inductance detector and some microwave microscope readout systems. To estimate this envelope-domain time-variation under the assumption of high resonant frequency, we make the *ansatz*, backed up by the simulation results, that the response voltage has the form:

$$\hat{x}(t) = [i(t) + jq(t)] \cdot e^{-j\omega_d t} = z(t) \cdot \tilde{a}_j e^{-j\omega_d t}, \qquad 3.69$$

where $\tilde{a}_j$ is the baseband excitation amplitude. Substituting this into the general parametric equation gives a second differential equation in terms of the complex time-varying envelope:

$$\ddot{z}(t) = \big(j2\omega_d - 2\sigma_1(t)\big)\dot{z}(t) + \big(\omega_d^2 - \omega_1^2(t) + j2\sigma_1(t)\omega_d\big)z(t) + \tilde{a}_j\omega_1^2(t) \qquad 3.70$$

Again this can be represented as a 2D 1st order ODE system through the introduction of the 'dummy' variables $z_2 = z(t)$ and $z_1 = \dot{z}(t)$:

$$\begin{aligned} \dot{z}_1 &= \big(j2\omega_d - 2\sigma_1(t)\big)\, z_1 + \big(\omega_d^2 - \omega_1^2(t) + j2\sigma_1(t)\omega_d\big)z_2 + \tilde{a}_j\omega_1^2(t) \\ \dot{z}_2 &= z_1 \end{aligned} \qquad 3.71$$

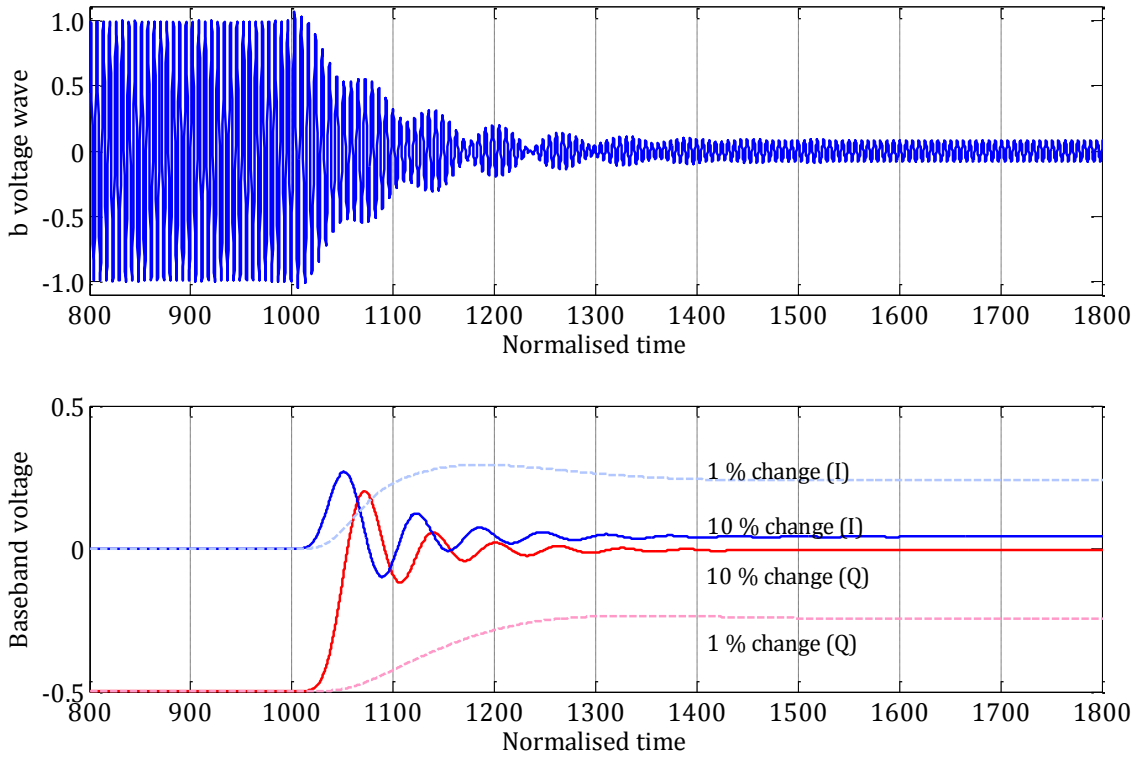



*Figure 3.13: Simulation of a transient change in resonator parameters. Top figure is the b voltage wave against time normalised to 1 Hz frequency. Lower figure is simulated baseband IQ signals; red is Q(t) and blue is I(t). A change in parameters of 1% in both σ and ω is shown as a continuous line, 10% change broken.*

The simulated response of this baseband 1st order ODE is shown above. The same delayed step response at baseband is closely approximated by:

$$x(t) \cong c_2 + (c_1 - c_2)\exp\{(\delta_\sigma - \sigma_1 + j\delta_\omega)(t-\tau)\}, \qquad 3.72$$

where $c_1$ and $c_2$ are the steady-state baseband responses for the initial and final conditions, respectively.

The 1st order ODE system can be used to construct another state-space representation, this time in terms of the 'baseband', where $\mathbf{z}(t) = [z_1, z_2]^{\mathrm{T}}$ is the state vector, and $\tilde{b}_i(t)$ is the baseband response:

$$\begin{aligned}\dot{\mathbf{z}}(t) &= \begin{bmatrix} j2\omega_d - 2\sigma_1(t) & \omega_d^2 - \omega_1^2(t) + j2\sigma_1(t)\omega_d \\ 1 & 0 \end{bmatrix}\mathbf{z}(t) + \begin{bmatrix} \omega_1^2(t) \\ 0 \end{bmatrix}\tilde{a}_j(t) \\ \tilde{b}_i(t) &= [K_1 \quad 0]\,\mathbf{z}(t) + K_0\tilde{a}_j(t)\end{aligned} \qquad 3.73$$

Although useful for simulation, this equation is still intractable as it stands. It therefore does not offer a clear route to estimating $\sigma_1(t)$ and $\omega_1(t)$. One way to incorporate the parameters into this equation is to include them in the state vector. This allows numerical methods that estimate the state vector, such as the Kalman filter [22], to estimate then the parameters at the same time. A deterministic model incorporating time-varying parameters is then be given by:

$$\begin{aligned}\frac{d}{dt}\begin{bmatrix} z_1 \\ z_2 \\ \omega_1 \\ \sigma_1 \end{bmatrix} &= \begin{bmatrix} (j2\omega_d - 2\sigma_1)z_1 + \left(\omega_d^2 - \omega_1^2 + j2\sigma_1\omega_d\right)z_2 + \omega_1^2\tilde{a}_j \\ z_1 \\ \dfrac{df_1(\boldsymbol{\theta},\omega_1,\sigma_1)}{d\boldsymbol{\theta}}\dfrac{d\boldsymbol{\theta}}{dt} \\ \dfrac{df_2(\boldsymbol{\theta},\omega_1,\sigma_1)}{d\boldsymbol{\theta}}\dfrac{d\boldsymbol{\theta}}{dt} \end{bmatrix} \\ \tilde{b}_i &= [K_1 \quad 0 \quad 0 \quad 0]\begin{bmatrix} z_1 \\ z_2 \\ \omega_1 \\ \sigma_1 \end{bmatrix} + K_0\tilde{a}_j(t)\end{aligned} \qquad 3.74$$

This can be seen to conform to a general state space model which is nonlinear only in the process equations (in a linear system $\mathbf{F}\big(\mathbf{w}(t),\mathbf{u}(t)\big) = \mathbf{A}\mathbf{w}(t) + \mathrm{B}\mathbf{u}(t)$):

$$\begin{aligned}\dot{\mathbf{w}}(t) &= \mathbf{F}(\mathbf{w}(t),\mathbf{u}(t)) \\ \mathbf{y}(t) &= \mathrm{C}\mathbf{w}(t) + \mathrm{D}\mathbf{u}(t)\end{aligned} \qquad 3.75$$

This state-space representation now incorporates the measurand vector $\boldsymbol{\theta}(t)$ as part of the excitation, or control input $\mathbf{u}(t) = \left[\tilde{a}_j(t), \theta_1(t), \theta_1(t)\right]^T$ through the use of the separated perturbation equations $f_1(\boldsymbol{\theta},\omega_1,\sigma_1)$ and $f_2(\boldsymbol{\theta},\omega_1,\sigma_1)$ – Equation 3.23 above, for example. The chain rule has been applied to these expressions to incorporate them in the differential process equation. It is an all-encompassing description of the resonator multi-stage model, which although useful for simulation purposes is impractical for use in estimation of real-time resonator parameters.

In order to find a more practical model, the nonlinear and dynamic aspects of the response can be simplified in an effort to separate them. The simulation results showed that a step-change in material properties could be accounted for by decomposing the response into a transient and steady state response. A simple equation for this baseband steady state response can be derived from the single pole model equations by substituting $\sigma_1 = -\omega_1/2Q_1$:

$$\frac{\tilde{b}_i(\omega_d,t)}{\tilde{a}_j(\omega_d,t)} = \tilde{s}_{ij}(\omega_d,t), \qquad \lim_{t\to\infty} \tilde{s}_{ij}(\omega_d,t) = K_0 + \frac{K_1(\omega_1+\delta_\omega)}{2(\sigma_1+\delta_\sigma)+j2(\omega_d-\omega_1-\delta_\omega)} \qquad 3.76$$

Thus for sufficiently slow changes in input parameter – *i.e.,* having a frequency content much less than the resonator bandwidth – this steady state solution provides a convenient form for the final stage of the resonator model proposed earlier. Thus, the change in the resonator spectral response function $\mathbf{S}(\omega,t) = g(\boldsymbol{p},t)$ can be written:

$$\Delta S_{ij}(\omega,t) = \left[S_{ij}(\omega,t) - S_{ij}(\omega,0)\right] = \frac{K_{1,ij}(\omega_1+\delta_\omega(t))}{2(\sigma_1+\delta_\sigma(t))+j2(\omega-\omega_1-\delta_\omega(t))} \qquad 3.77$$

Surface plots for this steady state change in spectral response function are shown below.

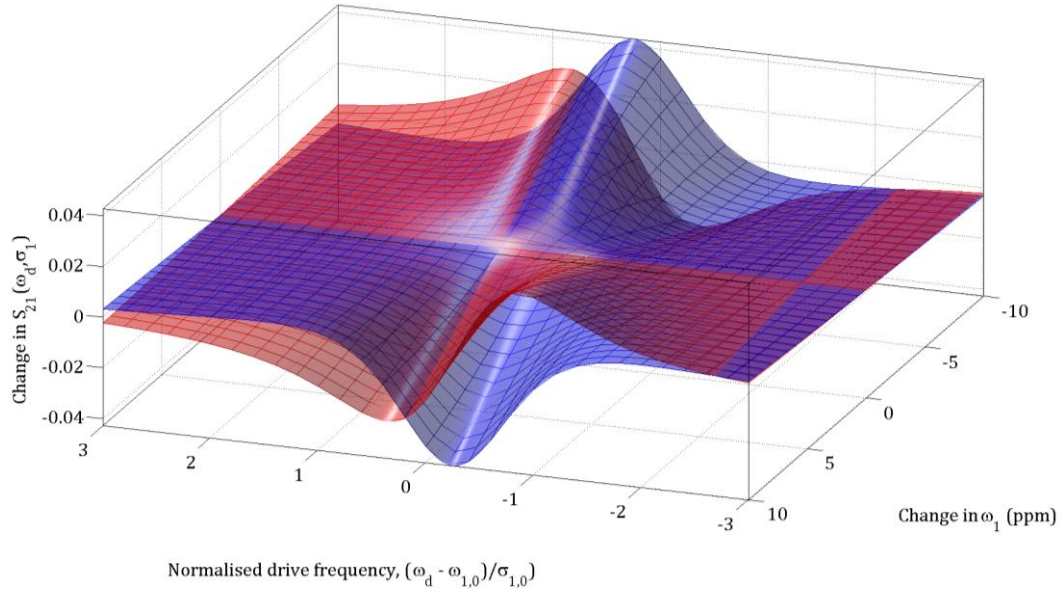



*Figure 3.14: Steady state change in scattering parameter for a small change in resonant frequency, blue surface is* $\Im\{\Delta S_{ij}(\omega,t)\}$ *and red surface is* $\Re\{\Delta S_{ij}(\omega,t)\}$*. Other values are* $Q_1(0) = 1000,\ K_1 = \frac{g}{1+g}\frac{1}{Q_1(0)}, g = 0.1$.

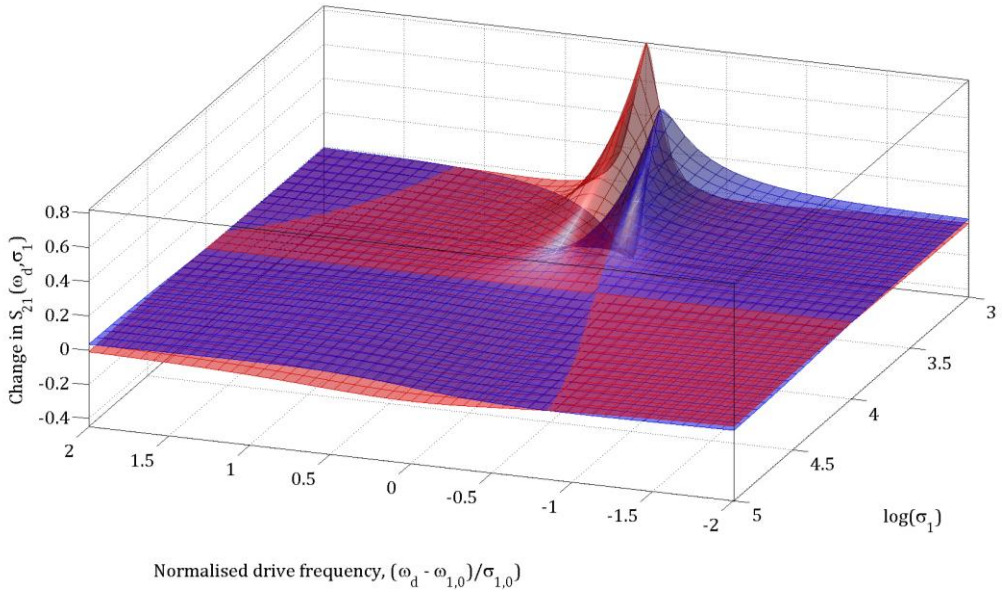



*Figure 3.15: Steady state change in scattering parameter for differing bandwidths. Surfaces and other parameters as above.*

This is the exact resonator perturbation equation for 'DC' input signals, and it shows that the response of the resonator to a change in its parameter is nonlinear. This means that if a parameter varies harmonically, $\omega_1(t) = \omega_1(0) + \delta_\omega e^{j\Omega t}$, even if slowly, then the output will consist of an infinite number of harmonics of this

signal: $\tilde{s}(t) = \tilde{s}(0) + s_1 e^{j\Omega t} + s_2 e^{j2\Omega t} + \cdots$ . Furthermore, the resonator response will not change instantly, even with a small perturbation, as has been demonstrated by the numerical simulation of the step response. Instead it will exhibit a response that varies both in amplitude and phase relative to the harmonically varying parameter, characterised by a time constant $\tau = 1/\sigma_1 = 2Q_1/\omega_1$ .

This composite nonlinear and dispersive response of the baseband is demonstrated by numerical simulation of the state space model (Figure 3.16) which shows the magnitude and phase response of the first three harmonics of the perturbation signal that would be received in I and Q baseband signals. This simulation shows that as the frequency of sinusoidal variation in the resonant frequency parameter increases, the measured response of the resonator decreases. If unaccounted for, this would lead to an error in any derived measurements. The values used in this simulation were: $\omega_1(0) = 10\ Grad/s$, $Q_1 = 500$, and $\delta_\omega = 1$ MHz which can be seen to give a first-order system type response in the first harmonic of $\Re\{z_1\}$ with cut-off $f_c \cong Q_1/\pi\omega_1 = 1.6$ MHz. First order behaviour is also seen in the roll-off of the *second* harmonic of $\Re\{z_1\}$; *i.e.,* -20 dB/dec, whereas the second harmonic of $\Im\{z_1\}$ rolls off at -40 dB/dec, and the third harmonic of $\Re\{z_1\}$, -60 dB/dec. However, these orders are inconsistent with the observed phase responses, so in general a first-order system model is not applicable to the harmonic distortion products.

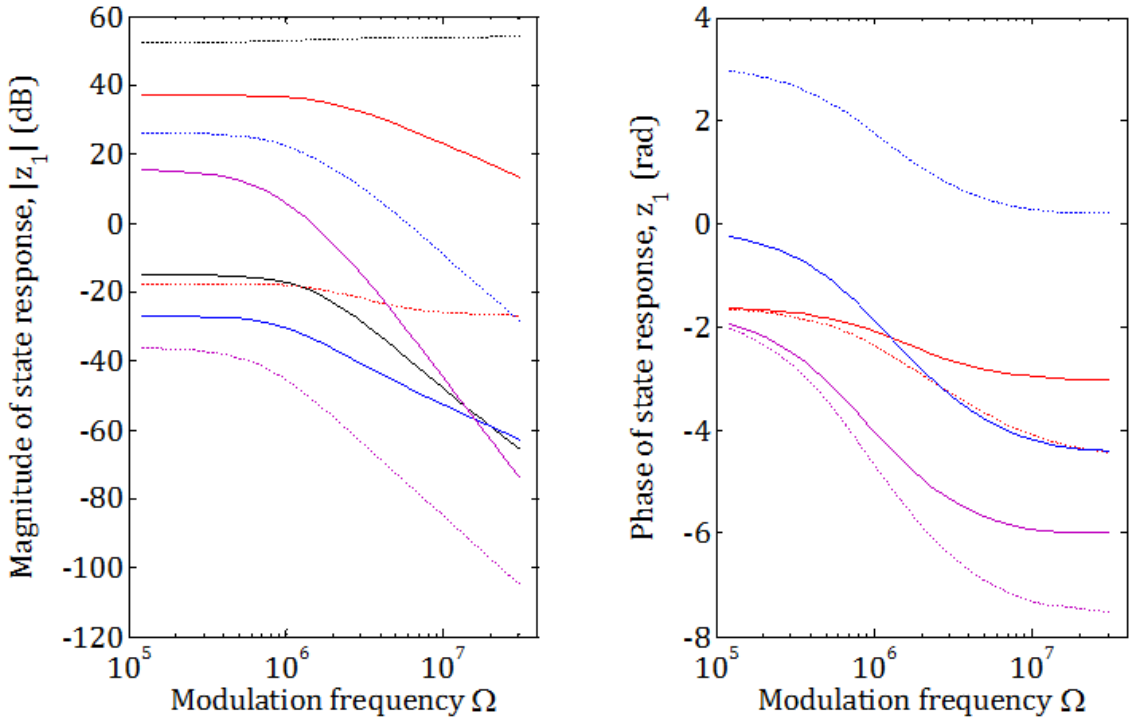



*Figure 3.16: Response of the resonator state variable, $z_1 = \tilde{b}_i/K_1$, to rapid sinusoidal variation in resonant frequency: $\omega_1(t) = \omega_1(0) + \delta_\omega sin(\Omega t)$. DC response is shown in black; first, second and third harmonics are shown in red, blue and purple, respectively. $\Re\{z_1\}$ is denoted by a continuous line and $\Im\{z_1\}$ is shown as a broken line. The magnitude (left) and phase (right) of the harmonics present in the state vector were obtained by Fourier analysis of the simulated time-domain response.*

In order to estimate the 'DC' response for the various harmonics, a multivariate Taylor series expansion of the response of the resonator to a small change $[\delta_\omega, \delta_\sigma]$ in its eigenvalues, shown here up to second order:

$$K_0 + \frac{K_1\omega_1}{2\sigma_1 + j2(\omega_d - \omega_1)} + \frac{K_1\big(\sigma_1 + j(\omega_d - 2\omega_1)\big)}{2\big(\sigma_1 + j(\omega_d - \omega_1)\big)^2}\delta_\omega - \frac{K_1\omega_1}{2\big(\sigma_1 + j(\omega_d - \omega_1)\big)^2}\delta_\sigma - \frac{K_1\big(\sigma_1 + j(\omega_d - 3\omega_1)\big)}{2\big(\sigma_1 + j(\omega_d - \omega_1)\big)^3}\delta_\omega\delta_\sigma - \frac{jK_1\big(\sigma_1 + j(\omega_d - 2\omega_1)\big)}{2\big(\sigma_1 + j(\omega_d - \omega_1)\big)^3}\delta_\omega^2 + \frac{K_1\omega_1}{2\big(\sigma_1 + j(\omega_d - \omega_1)\big)^3}\delta_\sigma^2 \qquad 3.78$$

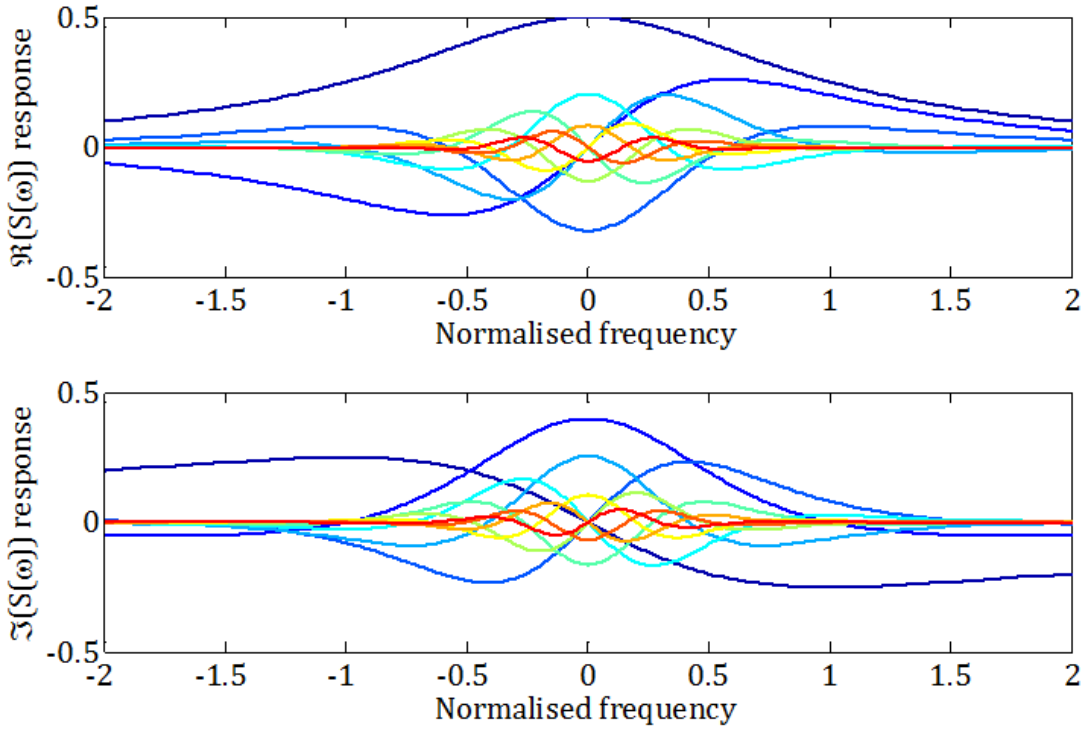



*Figure 3.17: First ten harmonics of the response against normalised frequency to a pure change in resonant frequency* $\delta_\omega = 0.8\,\sigma_1$*; DC – 10th harmonic are shown in colour spectrum progression blue – red.*

Combining this small signal, steady state approximation, with the first order dynamic behaviour observed it is possible to introduce a 2D frequency representation of the resonator. This gives the steady state response to a small harmonic variation in input parameter; a simple model for a lock-in system where material parameters are modulated by an external control. This would be the case, *e.g.,* when as a low frequency 'bias' field is applied in the case of ferromagnetic or ferroelectric materials, and that frequency is then demodulated by the MRS system receiver. As numerical simulation confirms, both in step response and harmonic response, the linear dynamic response can be modelled as a first order system with a single complex pole:

$$H(s') \sim \frac{(c_1 - c_2)(\sigma_1 - \delta_\sigma)}{s' + \sigma_1 - \delta_\sigma - j\delta_\omega} \xrightarrow{s'=j\Omega} H(\Omega) \sim \frac{(c_1 - c_2)}{1 + j\dfrac{\Omega - \delta_\omega}{\sigma_1 - \delta_\sigma}} \qquad 3.79$$

The complex baseband response therefore has two frequency dimensions, driving frequency and signal frequency. In *both* these dimensions, the resonator has a corresponding amplitude and phase, or equivalently real and imaginary, response. This is a 2D response over a 2D grid, impossible to visualise in one figure.

Combining the first order dispersion model with the first harmonic steady state response results in a multiple input, single output (MISO) system model parameterised over these two dimensions, 'slow' signal frequency $\Omega$ and 'fast' drive frequency $\omega$:

$$S_{ij}(\omega,\Omega) = \frac{K_{1,ij}}{1 + j\dfrac{\Omega - \delta_\omega}{\sigma_1 - \delta_\sigma}} \cdot \frac{\big(\sigma_1 + j(\omega - 2\omega_1)\big)\delta_\omega - \omega_1\delta_\sigma}{2\big(\sigma_1 + j(\omega - \omega_1)\big)^2} \qquad 3.80$$

This relatively simple function, technically a *describing function* as the first-order dispersion is dependent on both the frequency *and* magnitudes $\delta_\omega, \delta_\sigma$ of its two inputs, gives the *first harmonic* response of the resonator to a sinusoidally varying eigenvalue signal. The Taylor series can also be used to derive the transfer functions between the applied signal and its harmonics, if necessary. However, the resonator response to a more complex excitation signal consisting of multiple frequency components will be complicated by the mixing of its frequency content by the nonlinearity.

One approach to allow the response to arbitrary input signals to be estimated without resorting to time-domain simulation is to decompose the above model into a cascade of a linear dispersive system and a memoryless nonlinearity. This is known as a Wiener model when the nonlinearity is after the linear dispersion and a Hammerstein model when the nonlinearity is before it [23]. This requires the assumption that $\delta_\omega \ll \Omega$ and $\delta_\sigma \ll \sigma_1$ in order to remove the amplitude dependence from the dispersion model, making it a valid linear system. This unfortunately limits applicability to one of two limits: lower frequency and large amplitude or high frequency and small amplitudes; however it is more accurate than using either the steady state response or first harmonic approximation alone. From inspection of the simulated results, the linear dispersion should come before

the memoryless nonlinearity in order to produce the steeper roll-off seen in the higher harmonics (*i.e.,* a polynomial nonlinearity following a first order dispersion). Thus, an appropriate decomposition is the Wiener model:

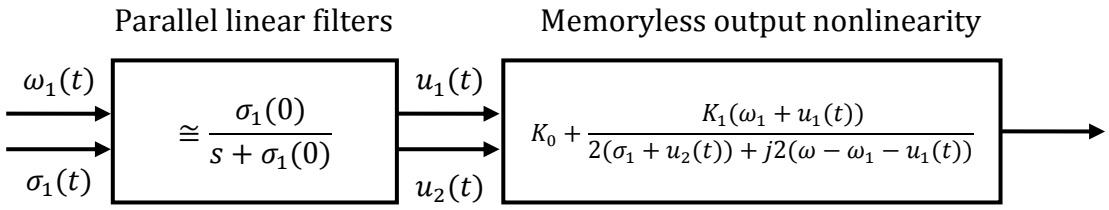



*Figure 3.18: Wiener model of the resonator spectral response function, valid for rapid perturbations of small amplitude, or low frequency perturbations of large amplitude.*

It is possible to find an inverse of this decomposed Wiener model; given a measurement of $S(\omega,t)$**,** pass this through the inverse nonlinearity $f^{-1}(S(\omega,t))$ and then filter it through the equalising filter $H^{-1}(s)$, which is a simple high pass filter in this case. To find the inverse nonlinearity, define two separate variables $i(t) \triangleq \Re\{\hat{s}(\omega,t)\}, q(t) \triangleq \Im\{\hat{s}(\omega,t)\}$ and solve simultaneously, which assuming the data have been normalised by $\hat{s}(\omega,t) = [S(\omega,t) - K_0]/K_1$ gives:

$$u_1(t) = \frac{2(\omega-\omega_1)[i(t)^2+q(t)^2] + \omega_1 q(t)}{2[i(t)^2+q(t)^2] - q(t)} = g_1(i(t),q(t)),$$
$$u_2(t) = \frac{\omega i(t) + \sigma_1(q(t) - 2[i(t)^2+q(t)^2])}{2[i(t)^2+q(t)^2] - q(t)} = g_2(i(t),q(t)). \quad 3.81$$

There is only one solution for this inverse so the resonator response is, perhaps surprisingly, unambiguous with respect to simultaneous changes in resonant frequency and bandwidth at any excitation frequency. To implement this inverse with a spectral measurement, the output could be simply summed over all the drive frequencies measured; *e.g.,* $\hat{u}_1(t) = \sum_\omega g_1(i(t),q(t))$. Although seemingly a good solution; mathematically an optimum estimator, in practice the 'unperturbed resonator' constants $\omega_1$, $\sigma_1$, $K_0$ and $K_1$ required to evaluate this inverse function are generally not known in advance, and there is a fundamental problem in defining and estimating the 'initial' or 'unperturbed' response $S(\omega,0)$ in an online

MRS system. Thus, in practice other methods such as curve fitting or neural networks must be used alongside or instead of this direct solution. Various methods for estimating and inverting the resonator dynamic spectral response, as well as a comparison of their performance in the estimation task and suitability for realtime computation will be discussed in Chapter 5. Note that, if curve fitting or similar is used to estimate the signal and resonator parameters, then the accurate inverse dispersion function $H^{-1}(s, \delta_\omega, \delta_\sigma)$ can be used to compensate for the roll-off of the resonator response – combined with a suitable envelope method (such as a Hilbert transform filter [24]) to estimate the amplitude, this should result in a good approximation.

This concludes the discussion of the modelling of resonator dynamic properties. Although modelling resonator behaviour through numerical integration of the continuous-time nonlinear state space equations is very accurate, it is also time consuming and computationally intensive (computing the 31 individual simulations necessary for Figure 3.16 took over 12 hours on a high-end workstation). The state space equations are also inflexible when modelling imperfections in the scattering parameters of the resonator – to take into account connecting transmission lines and load impedance mismatches, for example. Thus, the Wiener model is a good compromise between flexibility, computational complexity and accuracy, which is valid for arbitrary input signals of a dynamic nature – in other words a model suitable for the dynamic applications of MRS discussed in Chapter 2.

## 3.5 Statistical considerations

The performance of a sensor can be classified according to various measures. These can be represented graphically for a hypothetic one-dimensional sensor:

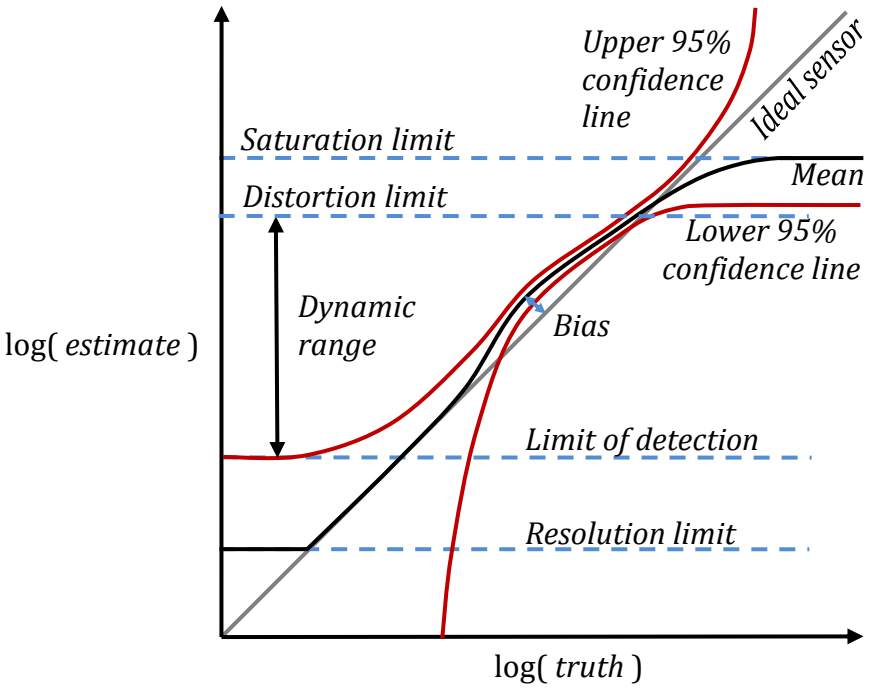



*Figure 3.19: Diagrammatic representation of general sensor response features (exaggerated to highlight key features).*

This diagram shows how several sources of error and uncertainty combine to affect the performance of a sensor. Fundamentally, from a statistical point of view, a sensor combines probability density functions from various error sources in order to produce a final, or posterior, probability density function which should be centred on a certain estimate value and reflect the degree of spread or distribution of the possible values that the estimate could be. This is represented in the above diagram by the use of central 'mean' line and a 'confidence interval' – the lines denoting the range where, given a large number of repeated trials, 95% of the measurements would be found to lie. It is possible to distinguish further two different sets of statistics in the sensor response – variation due to time-varying random errors and variation due to time-static systematic errors [14].

Random error determines the limit of detection and sets the bottom limit for the dynamic range parameter. This type of error is simpler to analyse and understand; each time a new measurement is made the 'error dice' are rolled again. Random error is due to noise. In passive microwave circuits, the main origin is Johnson-Nyquist noise caused by random motion of charge carriers having a kinetic energy proportional to temperature.

In a sensor system, there are also extrinsic and intrinsic noise sources. Extrinsic sources are those caused externally to the sensor, such as in the microwave electronics, connecting cables, *etc.,* and intrinsic sources are inherent to the sensor itself; in resonators, these are thermal fluctuations in the charge carriers in its conductors and lossy dielectrics. The effects of random fluctuations in magnetic and electric field energy can be estimated relatively simply – being 'small signal' characteristics, the effect of these noise sources can be 'propagated' through the resonator response model in order to determine their effect on the resonator.

Because such random errors are small, the propagation of the uncertainty due to their variation can be achieved by Taylor series expansion of the sensor response function to second order about the mean. To demonstrate this, consider the intrinsic noise sources in a lumped resonator (knowing that this can be generalised to a distributed resonator) [25]:

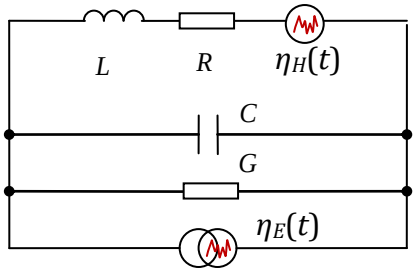



*Figure 3.20: Resonator noise sources – inductive branch noise voltage* $v_n(t) = \eta_E(t)$ *and capacitive branch noise current* $i_n(t) = \eta_H(t)$

The differential equations that describe the magnetic field and electric field branches of this noisy resonator can be expressed as first order stochastic equations, known as Langevin equations [26]:

$$L\frac{di(t)}{dt} = -Ri(t) + \eta_H(t) \qquad C\frac{dv(t)}{dt} = -Gv(t) + \eta_E(t), \tag{3.82}$$

where $\eta(t) = \mathcal{N}(0, \sigma^2)$ are zero mean random processes. To solve the Langevin equation for the distribution of voltage and current in these equations the first and second moments are calculated. Knowing that the equation is linear and the noise are zero-mean, it is clear that the mean voltage and current will not differ from their deterministic values; *i.e.,* $E\{i(t)\} = V_0 e^{-\frac{R}{L}t}$, and $E\{v(t)\} = I_0 e^{-\frac{G}{C}t}$. However, the variance, $E\{v(t)^2\} - E\{v(t)\}^2$, of the voltage and current is non-vanishing. The solutions for the variance are obtained using rules of statistical independence and the properties of differentiation in order to simplify the $E\{v(t)^2\}$ term [27]:

$$var\{i(t)\} = \left(\frac{\sigma^2}{2RL}\right)\left(1 - e^{-\frac{2R}{L}t}\right), \qquad var\{v(t)\} = \left(\frac{\sigma^2}{2GC}\right)\left(1 - e^{-\frac{2G}{C}t}\right) \tag{3.83}$$

According to the principle of equipartition of energy, the average energy of any passive harmonic oscillator is $k_B T$, divided equally between its kinetic and potential energy forms; *i.e.*, its magnetic and electric field energy. Knowing that the energy stored in the inductor is $\frac{1}{2}Li^2$ and the capacitor, $\frac{1}{2}Cv^2$, and taking thermal equilibrium to be the limit $t \to \infty$, substituting in the current and voltage variance, $i^2$ and $v^2$, gives these equilibrium equations:

$$\frac{1}{2}L\left(\frac{\sigma^2}{2RL}\right)=\frac{k_BT}{2} \quad \Rightarrow \quad \sigma^2=2k_BTR\equiv 2k_BT\frac{Y_R}{Q_c},$$
$$\frac{1}{2}C\left(\frac{\sigma^2}{2GC}\right)=\frac{k_BT}{2} \quad \Rightarrow \quad \sigma^2=2k_BTG\equiv 2k_BT\frac{Z_R}{Q_d}, \qquad 3.84$$

where $Q_c$ and $Q_d$ are the conductor quality factor and dielectric quality factor, respectively, and $Z_R=1/Y_R$ is the resonator impedance. Therefore, the intrinsic noise caused by the losses in the resonator is independent of the value of stored energy and therefore the resonant frequency. However, the spectral distribution of the noise measured across the whole resonator will not be white but will be 'filtered' by the resonator spectral response, as represented by the circuits below:

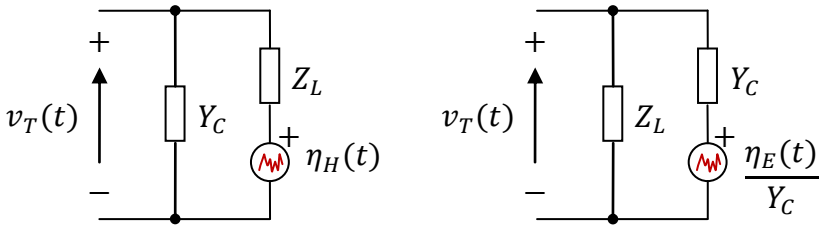



*Figure 3.21: Equivalent circuits showing the relationship between the individual branch voltage noise sources and the total voltage measured across the resonator.*

Because the individual noise sources are uncorrelated (valid in the case of thermal noise), the total noise variance measured across the resonator is simply the summation of the individual noise variances (equivalent to the combination of the noise voltages in quadrature). Furthermore, since the variance of a voltage is a measure of power and the power spectral density of noise passed through a linear filter is simply $S_{out}(\omega)=S_v(\omega)|H(\omega)|^2$ [28], the power spectral density of the noise measured across the resonator is given by:

$$var\{V_T(\omega)\}=\frac{2k_BT}{R}\left|\frac{1}{1+Z_LY_C}\right|^2+\frac{2k_BT}{G}\left|\frac{Z_L}{1+Z_LY_C}\right|^2=2k_BT(R+G)\frac{1+|Z_L|^2}{|1+Z_LY_C|^2}. \qquad 3.85$$

Expressed in terms of non-lumped element resonator equivalents, this Lorentzian-like power response is:

$$var\{V_T(\omega)\} = \frac{\left(\frac{Y_R}{Q_c} + \frac{Z_R}{Q_d}\right)\left(1 + \frac{Z_R^2}{Q_c^2} + \frac{\omega^2 Z_R^2}{\omega_0^2}\right)}{1 + \frac{2}{Q_c Q_d} + \frac{1}{Q_c^2 Q_d^2} + \frac{\omega^2}{\omega_0^2}\left(\frac{1}{Q_c^2} + \frac{1}{Q_d^2}\right) + \left(\frac{\omega^2 - \omega_0^2}{\omega_0^2}\right)^2}. \qquad 3.86$$

This function reveals that the peak voltage noise measured at the resonant frequency increases with Q factor, yet the total voltage noise integrated over all frequency stays constant – regardless of Q factor. This is simply because, on resonance, the resonator has increasingly high real impedance with Q factor. However, in all practical cases the thermal noise is below measurable limits. The peak voltage noise variance predicted for a room temperature resonator with $Q_c = 1000$, $Q_c \to \infty$ and $Z_R = 50$ is about 1 $\text{pV}^2$.

Whilst thermal noise generated by the loss mechanisms of a resonator may be too small to be of practical significance, there are other sources of noise that do dominate real resonators. Primary amongst these is the temperature-fluctuation induced change in the physical dimensions of a resonator. In a transmission line or cavity resonator, such tiny changes in the electrical length of the resonator cause a shift in the resonant frequency. Furthermore, the electrical permittivity and magnetic permeability of materials within the resonator is also temperature dependent; hence, temperature fluctuations can also cause the equivalent capacitance and inductance of the resonator to vary.

The situation is much more complex if the inductance and capacitance – or equivalently the resonant frequency and bandwidth – are taken to be stochastic variables. The Langevin equation approach is too longwinded to present here – this approach is discussed at length in [29] for general second order systems, although the case where damping and resonant frequency are both random variables simultaneously is not considered. However, a simple static modelling approach can be used to estimate the magnitude of such small fluctuations in resonator parameters on the resonator spectral response. In fact, the response of

the resonator to such fluctuations is the same (and hence inseparable from) the small-signal response of the resonator to a desired signal, as previously derived in Equation 3.78. Thus, the spectral response variance arising from a small fluctuation in resonant frequency and bandwidth is given by:

$$var\{S_{ij}(\omega_d)\} = \frac{|K_1|^2}{4} \cdot \frac{\left([\sigma_1^2 + \omega_1^2] \cdot \sigma_{\omega_1}^2 + \omega_1^2 \cdot \sigma_{\sigma_1}^2 - 2\omega_1\sigma_1 \cdot \sigma_{\omega_1\sigma_1}\right)}{\left|\left(\sigma_1 + j(\omega_d - \omega_1)\right)^2\right|^2}. \qquad 3.87$$

In this equation, $\sigma_{\omega_1}^2$ is the variance of fluctuations in resonant frequency, $\sigma_{\sigma_1}^2$ is the variance of fluctuations in bandwidth, and $\sigma_{\omega_1\sigma_1}$ is the covariance of fluctuations between resonant frequency and bandwidth. If there is a single perturbing noise factor, such as temperature fluctuations, that affects both resonant frequency and bandwidth, these (co)variances can be linearly related to it under the small signal approximation:

$$\sigma_{\omega_1}^2 = \left(\frac{\partial\omega_1}{\partial\eta}\right)^2 \cdot \sigma_\eta^2, \quad \sigma_{\sigma_1}^2 = \left(\frac{\partial\sigma_1}{\partial\eta}\right)^2 \cdot \sigma_\eta^2, \quad \sigma_{\omega_1\sigma_1} = \frac{\partial\omega_1}{\partial\eta}\frac{\partial\sigma_1}{\partial\eta} \cdot \sigma_\eta^2 \qquad 3.88$$

In these expressions $\eta$ is the perturbing 'noise' variable, for example temperature, and the derivatives $\partial\omega_1/\partial\eta$ and $\partial\sigma_1/\partial\eta$ are the sensitivities of the resonant frequency and bandwidth to the noise variable. These equations can also be related to fluctuations in capacitance and inductance, where relevant.

In the case of thermal expansion of a cavity resonator, for example, these terms would be related to the thermal expansion coefficient of the resonator material, the dimensions of the resonator and the number of wavelengths, resonant frequency, and geometry of a particular mode. The effect of temperature on bandwidth would be insignificant in most cases.

Whilst this static approach is useful to estimate the magnitude of noise effects, comprehensive dynamic modelling of parametric noise can be accomplished if its probability distribution is known. This is useful when the distribution of noise is known to follow some specific spectral distribution, such as $1/f^n$. A comprehensive dynamic state-space noise model can be expressed as:

$$\begin{aligned}\dot{\mathbf{w}}(t) &= \mathbf{F}\big(\mathbf{w}(t), \mathbf{u}(t)\big) + \boldsymbol{\nu}_1(t) \\ \mathbf{y}(t) &= \mathrm{C}\mathbf{w}(t) + \mathrm{D}\mathbf{u}(t) + \boldsymbol{\nu}_2(t)\end{aligned} \qquad 3.89$$

This deceptively simple stochastic differential equation now includes the process noise $\boldsymbol{\nu}_1(t)$, which by virtue of incorporating the resonant frequency and bandwidth information in the state vector, now includes noise caused by fluctuations in capacitance and inductance, as well as thermal noise. This can model effects such as thermal instability in resonator geometry, mechanical vibration (*i.e.,* acoustic noise or microphonics), and more obscure phenomena such as the two level system fluctuations that cause capacitor noise in kinetic inductance detectors (see Section 2.3.1 for a discussion). The addition of measurement noise $\boldsymbol{\nu}_2(t)$ can further model mixer and amplifier noise as well as quantisation noise in analogue to digital conversion. The simple fact that the amplitude of the signals transmitted through the resonant system are dependent on the measured parameters further means that the signal to noise ratio of the output varies with input bandwidth. Thus resonator measurements inherently exhibit heteroscedasticity. This means that the variance of the output variables (*i.e.,* parameter estimates) changes with their value.

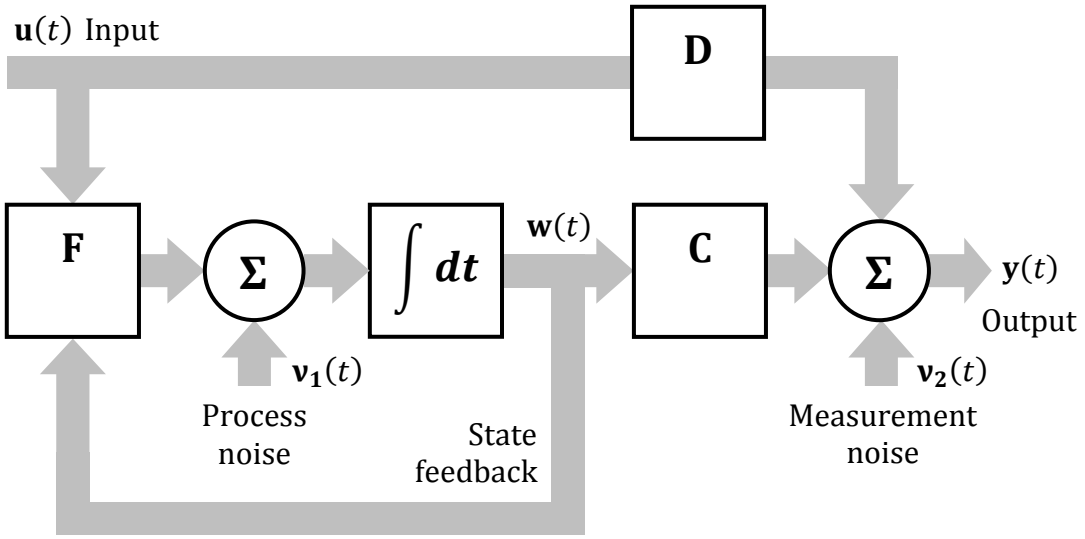



*Figure 3.22: Diagrammatic representation of the continuous-time stochastic state space model. After [22].*

Having established that the effects of random errors can be simulated effectively using the stochastic state space equations, the next and final discussion for this section refers the what is often the dominant source of error in microwave resonant sensors; 'static' or systematic error. This source of error is often seen as more pathological and more difficult to detect, quantify, and reduce. That is because, unlike random error, systemic error is the same each time the 'dice are rolled' on a new measurement. Thus, no amount of averaging or simple statistical methods can be used to reduce it in a given experimental setup.

There are two primary sources of systematic error; these are model parameter error and nonlinear distortion. Model parameter error occurs when we use a model of the system to produce a measurement but the parameters of that model, or even the model itself, are not true to life. In effect this leads to a permanent under- or over-estimate of the true measurement; known as bias. Model parameter error can arise because the measurements of the real life system are flawed – for example we may have measured the radius of a cavity resonator to be 10mm, but in fact it is actually 10.1mm. Using the same example, model error would arise if we assumed that the cavity resonator had a circular cross section when actually it

was ovoid. These types of error are universal to all components, *e.g.,* finite machining accuracy leading to tolerance in component parameters.

To alleviate this source of error one might decide to calibrate using a measurement of a known material. In theory this could be seen to eliminate systematic error (assuming a linear model!), as one could correct for gain errors with just one sample (and there is no offset error in a differential measurement). However, there might be uncertainty in the actual 'reference' value of the sample. The microwave resonator technique is already the most accurate way of measuring permittivity, and all manner of environmental and dimensional effects mean that a sample measured as a reference somewhere else may actually have a different permittivity in a current experiment [30].

This is especially true for the measurement of the permittivity of liquids and solids, and is confounded by the fact that, as yet, there is also no theoretical prediction for the permittivity of materials which can be used as a 'gold standard'. Rarefied gasses do have accurate theoretical permittivity values, so perhaps these are the way to calibrate. But this then becomes an exercise in accurate temperature and pressure measurement in order for the theoretical prediction to agree with reality, and the values of range of different permittivity values that could be calibrated in this way is very small. Furthermore, any calibration measurement made will itself be subject to random error. This will then be introduced into the calibrated system as a source of systematic error. Thus, calibration does not fully eliminate the issue of systematic error. I would go so far as to say that calibrating microwave resonant sensors with reference materials should be done with extreme caution, if at all. Due to the aforementioned problems with 'gold standards', however, this assertion is somewhat untestable at present.

Nonlinearity can also cause systematic error; in some ways, this can be seen as a type of model error. Failure to correct for nonlinear effects and distortion is akin to failure to have a sufficiently representative model. The ultimate nonlinearity is saturation and no amount of correction can account for this. In microwave resonators 'saturation' as such is difficult to define. If the resonance moves outside

a limited tuning range, this can be considered to be a type of saturation. In addition, if the bandwidth increases to the point where parameter estimate is saturated by noise (the heteroscedasticity of the measurement because of insertion loss increasing) then this can be considered too to be a form of saturation. The fact that these limits generally only occur when the sample perturbation is very large means that resonant sensor measurements are generally considered to have a high dynamic range.

However, nonlinearity can be more pervasive than simple saturation. Nonlinearity not only distorts a signal, but it also distorts random noise too. Thus, the statistics of a signal are altered by nonlinear systems, potentially leading to bias. Since bias does not reduce with averaging or filtering, it becomes a source of systematic error. Hence, nonlinearity turns random noise into systematic error. This may seem like a minor issue and in conventional VNA measurements, it probably is. However, the new digital radio architecture developed to replace VNA methods in this thesis necessarily does not have same dynamic range as the VNA for spectral response measurements. Fundamentally, this is because the VNA is a narrowband radio receiver, meaning it has a small equivalent noise bandwidth and can use a high resolution A/D converter with low quantisation noise. Because the software defined radio has a noise bandwidth equal or exceeding the resonator bandwidth, and must therefore also use a lower resolution, faster A/D converter, its dynamic range for spectral measurements is reduced.

In order to restore the 'lost' dynamic range of spectral measurements and obtain an equivalent precision for the resulting resonant sensor measurements, essentially some form of averaging must be used. In order to avoid systematic error, therefore, the minimisation of distortion-induced systematic error is vital to ensure absolute accuracy is not degraded. Thus, what is a minor issue becomes of critical importance for the success of this approach.

The following example should help to demonstrate why it is important to be able to correct nonlinearities at the full sample rate, and show that only linearisation of the data *before* averaging will result in the reduction of bias. This is a key point,

and can be easily overlooked; this example refers to a classic mistake in spectral measurements. Suppose an instrument outputs phase and magnitude information of a complex signal subject to some corruption by noise. Suppose also that the user wants the real and imaginary parts of the data, (the measurands $r$ and $i$) which they know can be easily extracted from this information. To improve SNR, it is decided to average the phase and magnitude results coming from the instrument (which could be also accompanied by sample rate reduction, or decimation). However, the resulting average, although it might appear to fluctuate less and be less 'noisy', actually leads to an increasingly poor estimate of the desired real and imaginary parts of the signal as averaging is increased.

The reason is the nonlinear transformation, *i.e.*, distortion, caused by the magnitude and phase transformations; $f_1(r,i) = \sqrt{r^2 + i^2}$, $f_2(r,i) = \tan^{-1} i/r$. If both $r$ and $i$, nominally constant, are assumed to be corrupted by uncorrelated noise; $r \sim \mathcal{N}(r_0, \sigma_r^2)$, $i \sim \mathcal{N}(i_0, \sigma_i^2)$ a direct average of the linear measurands should tend to the true value $r_0$ as $E\{r\} = r_0$ in the limit of infinite averages – the mean ergodic theorem [28]. However, it can be shown by Cramér's delta method [31], that expected value of a *nonlinear* function does not generally tend to the same value as it would in the presence of no noise – in other words it exhibits bias:

$$E\{f_1(r,i)\} \cong \sqrt{r_0^2 + i_0^2} + \frac{r_0^2\sigma_r^2}{2(r_0^2 + i_0^2)^{3/2}} + \frac{i_0^2\sigma_i^2}{2(r_0^2 + i_0^2)^{3/2}} = f_1(r_0, i_0) + bias. \qquad 3.90$$

Cramér's delta method is used here simply to indicate a general approximation valid for arbitrary nonlinear functions. Indeed, there is an exact equation for the probability density function of the magnitude of a complex variable, when its real and imaginary parts are zero mean, independent and identically distributed; the Rayleigh distribution, $p(x) = x\sigma^{-2}e^{-x^2/2\sigma^2}$, which has mean $\sigma\sqrt{\pi/2} \neq 0$. This is also a specific case of the noncentral Chi distribution for the norm of any number of independent normally distributed variables. Similarly, the expected phase is

$$E\{f_2(r,i)\} \cong \tan^{-1}\left(\frac{i_0}{r_0}\right) + \frac{r_0\sigma_r^2}{(r_0^2 + i_0^2)^2} - \frac{i_0\sigma_i^2}{(r_0^2 + i_0^2)^2} = f_2(r_0, i_0) + bias. \qquad 3.91$$

Attempting to apply the inverse transformation to recover the measurands $r$ and $i$,

$$r = [f_1(r_0, i_0) + bias]\cos(f_2(r_0, i_0) + bias),$$
$$i = [f_1(r_0, i_0) + bias]\sin(f_2(r_0, i_0) + bias), \quad 3.92$$

now clearly gives the wrong results. Even if the phase bias cancels when $\sigma_r^2 = \sigma_i^2$, (this arises because arctangent is an odd-symmetric function around the expected value) the magnitude does not have this specific symmetry and hence leads to an overestimate of the amplitudes of the measurands.

These results show that to avoid bias it is good practice to convert each measurement to one linearly related to the main source of noise before averaging. If the noise and the signal are made linearly separable – *i.e.*, in the form $s(t) + n(t)$ they must therefore be uncorrelated. Linearising therefore has the property of minimising correlation of noise with a signal – a fact exploited in a majority of blind linearisation (blind meaning without prior knowledge of the desired signal) and predistortion algorithms (predistortion means linearising by correcting the input to a nonlinear system rather than the output). Generally for a nonlinear function, $E\{f(\mathbf{x})\} \neq f(E\{\mathbf{x}\})$, therefore this principle of linearisation at maximum rate is key to getting accurate results when any form of averaging or sampling rate reduction of the output of nonlinear system is being used.

In the case of the resonator spectral measurements, which are nonlinear functions of $\sigma_1(t)$ and $\omega_1(t)$, if they are modelled as random variables through the process noise $\mathbf{v}_1(t)$, then averaging any spectral estimates before calculating $\sigma_1(t)$ and $\omega_1(t)$ directly may lead to bias. However, the measurement noise $\mathbf{v}_2(t)$ is linearly related to the output $\mathbf{y}(t)$, thus it is most effectively reduced by averaging the spectral estimates directly. Passing the noise due to $\mathbf{v}_2(t)$ through whatever nonlinear inverse function is being used to estimate $\sigma_1(t)$ and $\omega_1(t)$ will lead to bias! This is a tricky situation, and it means that if both the measurement noise and the process noise are significant, a minimum bias can only be achieved through a balance of averaging before nonlinear inversion and averaging after nonlinear inversion.

Once $\sigma_1(t)$ and $\omega_1(t)$ have been estimated, there is still the issue of the nonlinear resonator eigenvalue response function. For the simple static field case this is given by Equations 3.23 and 3.25. It is in the eigenvalue response function that the effects of systematic error are felt. Often the largest source of error is the uncertainty in the sample volume $V_S$. If, for simplicity, the field over the sample is assumed to be uniform, then

$$\frac{a}{c} = \frac{\int |E(0)|^2 dV_S}{2\int \varepsilon |E(0)|^2 dV_R} \cong V_S \frac{|E_0(0)|^2}{4U_t(0)}, \tag{3.93}$$

Although the sample is static, for the purposes of determining systematic error $a$ can be treated as a random variable $a \sim \mathcal{N}(\mu_a, \sigma_a^2)$ or some other distribution. Indeed it may be a composite of several other geometric parameters, such as length, diameter, etc., which may or may not be independent. In this case, the full covariance matrix of all significant geometric parameters must be taken into account. In the simple Gaussian case, propagation of one-variable error $E\{f(x)\} \cong f(E\{x\}) + \frac{1}{2} f''(E\{x\})\sigma_x^2$ and $\text{var}\{f(x)\} \cong f'(E\{x\})^2\sigma_x^2$ leads to:

$$\hat{\theta}_1 \sim \mathcal{N}\left(\theta_1\left(1 - \frac{\sigma_a^2}{a^2}\right), \quad \frac{\theta_1^2}{a^2}\sigma_a^2\right). \quad \hat{\theta}_2 \sim \mathcal{N}\left(\theta_2\left(1 + \frac{\sigma_a^2}{a^2}\right), \quad \frac{\theta_1^2}{a^2}\sigma_a^2\right) \tag{3.94}$$

Even in this simplest case, an uncertainty in sample volume can be seen to cause bias, although since $\sigma_a^2 \ll a^2$ this is small. If analytic equations do exist for determining resonator perturbation parameters from geometric uncertainty distributions, then the general second-order approximation for the first and second moments of a multivariate function (Cramér's delta rule applied to multiple dimensions) can be used:-

$$E\{f(\mathbf{x})\} \cong f(E\{\mathbf{x}\}) + tr\left\{\frac{1}{2}\mathbf{B}^T\boldsymbol{\Sigma}\right\}, \qquad where \;\; \mathbf{B} = \boldsymbol{\nabla}^2 f(E\{\mathbf{x}\})$$

$$\text{cov}\{f(\mathbf{x})\} \cong \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T, \qquad where \;\; \mathbf{A} = \boldsymbol{\nabla} f(E\{\mathbf{x}\}) \tag{3.95}$$

The vector $\mathbf{x}$ should contain all the random variables with known covariance matrix $\boldsymbol{\Sigma}$. In some literature, this is known as a sensitivity analysis.

In practice, however, an analytical solution for which the Jacobian and Hessian can be calculated is unlikely to exist in the case of determining the resonator perturbation parameters. Therefore a powerful alternative method is to use Monte Carlo methods combined with numerical or FEM simulation to determine the output spread. This involves drawing the random variables used for a simulation from a known distribution (with known correlations). Over many thousands of simulations, a distribution function of the output variables is created. This naive approach is very inefficient, and more optimal techniques are available that involve creating nonparametric models for the sensitivity of a simulation to its input variables.

Although uncertainty, noise and error analysis is often too complex to allow closed-form solutions, what this section has showed is that at the very least there are appropriate processes to simulate and estimate the dynamic performance of resonant sensors, and therefore provide for a means to compare and ultimately optimise a sensor system.

## 3.6 Solving the inverse problem

To determine the target parameter estimate, or measurands, $\widehat{\boldsymbol{\theta}}$ from the eigenvalue requires that an estimate of the inverse of $\mathbf{p}(t) = f(\boldsymbol{\theta}(t))$ be found – a procedure known as inverse modelling. Thus, the equation that is actually executed during a measurement is:-

$$\widehat{\boldsymbol{\theta}}(t) = f^{-1}\big(\mathbf{p}(t)\big). \tag{3.96}$$

The normal way to go about this is to first express the forward function analytically, and then find a mathematical inverse of it to make the measurement. This was shown above for the simple linear case:

$$\widehat{\boldsymbol{\theta}}(t) = \mathbf{K^{-1}}\frac{(\mathbf{p}(t) - \mathbf{p}_0)}{|\mathbf{p}(t)|^2}. \tag{3.97}$$

However, this way to estimate an inverse function raises a number of issues. The first and most important issue is that any analytical solution rests on simplifying assumptions, as analytical complexity can increase staggeringly fast. Consider the difference in the expressions expanded to first and second order in $\theta$ above (the second order equation has no analytical inverse, incidentally). Thus, it is usually not practical to express an analytical solution that is sufficiently true to real life, no matter how elegant its mathematical expression might be.

Another important issue in any multivariate system such as this is the number of degrees of freedom represented by its describing equations. A measurement of two independent parameters can be used to estimate at most two independent unknowns. Thus, even though the eigenvalue might depend on a very large number of parameters, only two can be found by the function inverse at any one time. That implies that all the other perturbing factors of the resonator must remain constant over the course of a measurement, or be determined separately using some other measurement.

A further issue is one of solution multiplicity. Only a monotonic function, *i.e.*, one that only increases (or decreases) with $\boldsymbol{\theta}$, will have only one solution when inverted. A simple example of a non-monotonic function is $y = x^2$, whose inverse, $x = \pm\sqrt{y}$, clearly has two solutions. This is bad news, as it results in ambiguity. Whilst a human might be able to infer or assume a certain solution from the context of the problem, such an inference can be difficult for a machine to do in an automated sensor system. Non-monotonic resonator responses can occur when sample depolarisation is significant as well as for unusually large filling factors, so this consideration does have practical implications. A final problem is in actually finding an inverse of a function at all, as some representations of equations (*e.g.*, polynomial expansions) can have intractable roots. This is less significant in practice, however, since numerical interpolation, or 'look up tables', can be used as long as the forward equation can be evaluated.

All these issues can be bypassed by expressing the forward model in a different way, one for which an inverse is much easier to find. In effect, the problem of inversion becomes a task of regression. Regression means fitting a function, whether to measured or simulated data, that aims to approximate the true inverse function well enough to be useful. Fortunately, some 'tricks' can be used to prop-up a predominantly simulation-based model by a scarcity of real data points. This ensures that even if available real-life calibration is limited it does not mean having to reduce the model to a cripplingly inaccurate one.

All smooth differentiable functions can be approximated by a finite polynomial via Taylor's theorem, and the above discussion demonstrated that the eigenvalue response function for a two-parameter input may be written:

$$\mathbf{p} = [\theta_1, \theta_2, \ \ \theta_1^2, \theta_1\theta_2, \theta_2^2, \ \ \theta_1^3, \theta_1^2\theta_2, \theta_2^2\theta_1, \theta_2^2, \ \ \ldots]\mathbf{A} + \mathbf{p}_0 \qquad 3.98$$

$\mathbf{A}$ is a $2 \times M$ matrix of weights, where $M = N(N+3)/2$ if N is the highest order of the expansion. Note the cross terms such as $\theta_1\theta_2$ in the expansion that are necessary for a fully expressive representation of multivariate equations. Such nonlinear mixing of the two variables produces nonlinear correlations between the

dimensions of $\mathbf{p}$ which cannot be rectified by linear transformation. This cross-term expansion is equivalent to the nested sequence $\mathbf{x} = [x_1, x_2, x_3, \dots]$ of convolution of the ordinary polynomial sequences along each dimension of $\boldsymbol{\theta}$, *i.e.,* $x_n = \{\theta_1^{n-k} \cdot \theta_2^k\}_{k=0}^n$. This representation is known as a general polynomial regression model, and the matrix $\mathbf{A}$ can be estimated from a series of $n$ simulation data, $\mathbf{p}_i = \mathbf{x}_i\mathbf{A} + \mathbf{p}_0$ $(i = 1 \dots n)$, by forming the matrix $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n]^T$ of observations, and the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$ of convolved polynomials in known values of $\boldsymbol{\theta}$ and using the least squares solution based on the Moore-Penrose pseudoinverse [32] to estimate the weight matrix:

$$\widehat{\mathbf{A}} = (\mathbf{X}^T\mathbf{X})^{-\mathbf{1}}\mathbf{X}^T(\mathbf{P} - \boldsymbol{I}\mathbf{p}_0). \qquad 3.99$$

This provides a reasonably robust and reliable estimate of the forward model in the absence of noise, which matches closely simulations of resonators having samples in weakly depolarising regimes. However, the overall function in $\boldsymbol{\theta}$ does not have a defined inverse in general and an inverse must be found numerically. If the transformation $\mathbf{x}(\boldsymbol{\theta})\mathbf{A}$ has an inverse, for example $\mathbf{x}(\cdot)\mathbf{B}$ in the same notation, so that $\mathbf{x}(\,\mathbf{x}(\boldsymbol{\theta})\mathbf{A}\,)\mathbf{B} \cong \boldsymbol{\theta}$, then that function could be performed on the eigenvalue estimates on a measurement-by-measurement basis in order to extract the measurands in real time. Since low-dimensional matrix multiplications are relatively fast computationally, such an operation could be performed by a real time system in principle, allowing the nonlinearities of a resonator to be corrected at the same rate as samples are collected.

Having established the importance of linearisation, we are left with the problem of finding an appropriate inverse function. Although polynomial representations arise naturally from Taylor expansions, as in the extended resonator perturbation theory shown earlier, this representation isn't a good one for producing a reliable inverse as their solution are intractable at higher orders and they suffer from a multiplicity of solutions. A multivariate polynomial expansion can be written $\mathbf{x}(\boldsymbol{\theta})\mathbf{A}$ – this approximates a function as a weighted sum of nonlinear *basis functions*. The function $\mathbf{x}(\boldsymbol{\theta})$ is known as the basis generating function. In the above example, this function generated the polynomial cross products of its input

dimensions. But polynomials are just one family of such nonlinear basis functions, and some are perhaps better suited to the problem in hand. The key to making this move to a more appropriate approximation method is that the choice of basis function doesn't necessarily have to have any physical meaning in order to do the task in hand arbitrarily well. In other words, we move into the realm of nonparametric modelling, where the model does not necessarily have to have physical meaning, but merely has to approximate some data to a required degree.

### 3.6.1 Basis functions and Artificial Neural Networks

One type of basis function that is commonly encountered for approximating a function is the Fourier basis. The basis functions here are sines and cosines of different frequencies and the Fourier series is a weighted sum of these functions – the weights are thus the coefficients found by performing the Fourier transform. Whether or not the Fourier series is a good approximation depends on the type of function. Those having discontinuities tend to be less well approximated in the locality of the discontinuity. This is known as Gibbs' phenomenon. Other basis functions have similar issues. For example, Runge's phenomenon occurs at the edges of a polynomial basis approximation leading to local non-convergence of the approximation with increasing order [33]. One important class of approximants are the radial basis functions (RBF). Each output from an RBF machine is given by:

$$y(\mathbf{x}) = \mathbf{w}^T \phi(\|\mathbf{x} - \mathbf{x}_0\|) = \sum_{k=0}^{n} w_k \phi_k \left(\|\mathbf{x} - \mathbf{x}_0\|\right) \qquad 3.100$$

This is a sum of distance functions, weighted by the vector $\mathbf{w}$. The term $\|\mathbf{x} - \mathbf{x}_0\|$ is a distance norm, often the Euclidean distance, from some centre vector $\mathbf{x}_0$ – note the similarity here to the multivariate Taylor expansion about a particular point. The nice feature of RBF approximation is that, given a known or fixed centre vector, the weights can be estimated from 'training' data using one of the many methods for solving linear equations, such as the pseudoinverse. Indeed, if there are as many centres as training data points and these centres are taken to be equal to the $\mathbf{x}_i$ values of the training set, then the matrix of distances from each $\mathbf{x}_i$ to

these centres, $\mathbf{\Phi}$, is a square, positive definite matrix having a definite inverse, thus $\hat{\mathbf{w}} = \mathbf{\Phi}^{-1}\mathbf{y}$. This is known as the RBF interpolation, as the function it implements must pass exactly through its training points.

Typical basis functions used are Gaussian, $\phi(r) = e^{-\beta_k r^2}$ and inverse quadratic (*i.e.,* Lorentzian), $\phi(r) = 1/[1 + (\beta_k r)^2]$. The parameter $\beta_k$, or shape parameter, determines the selectivity or 'bandwidth' of the function, with smaller values tending to increased flatness. Whilst this parameter can be tuned individually, it is often fixed to a constant value for all the basis functions in order to simplify the estimation. If this is the case, it is important to normalise the input data so that the expected variance of all input dimensions is approximately equal. Indeed, this is good practice for most non-parametric methods of function approximation. From a small simulation set this is done simply by normalising by the nominal range of each dimension; a large measured dataset could instead be normalised by prewhitening using its estimated covariance matrix. This would further enhance performance by linearly decorrelating the input dimensions before the RBF approximant.

Closely related to RBF machines are Artificial Neural Networks (ANN). In fact, an RBF can be viewed as a simple type of ANN. The difference between ANNs and the other types of basis function approximant previously mentioned is simply the location of the linear weights. Whereas in basis function approximants the linear weighting and summation occurs after a nonlinear function, in an ANN the linear weighting and summation occurs *before* the nonlinear function. This subtlety can significantly enhance the expressive power of ANNs, making them more efficient at learning arbitrary functions. However, this complicates the estimation of their weights, as linear methods such as the pseudoinverse are no longer applicable. In a fanciful analogy to biological neural networks, these weights are called 'synaptic weights' and the nonlinear functions 'activation functions'. Fortunately, any equivalence to real biological neural networks has long been dismissed [34]; this should not detract from these otherwise extremely powerful and general tools in the field of machine learning. Each output from an ANN, with a single layer of nonlinear neurons of activation function $f(\cdot)$ is given by:

$$y(\mathbf{x}) = f(\mathbf{w}^T\mathbf{x} + x_0) = f\left(\sum_{k=0}^{n} w_k x_k + x_0\right) \qquad 3.101$$

The bias scalar, $x_0$ is usually neglected under the assumption that it is included in the input, *i.e.,* $\mathbf{x} = [x_1, x_2, \dots, x_d] \rightarrow [1, x_1, x_2, \dots, x_d]$. The weight for this 'zeroth' input thus becomes the bias. Although useful, this current ANN has only one layer – only one set of weights and activation functions. This network is not capable of learning arbitrary functions, and in fact implements what is known as a linear discriminant function. This is a bounded output than can loosely be interpreted as a probability; in other words a form of binary classifier. To make a neural network capable of regression (and interpolation), it is necessary to add another output stage which can linearly weight the outputs of this layer. As shown by Kolmogorov, a neural network implementing a weighted linear combination of an unlimited number of arbitrary nonlinear neurons, *i.e.,* $f(\mathbf{x}) = \sum_j \Xi_j \left(\sum_i \psi_{ij}(x_i)\right)$, is capable of *universal approximation* [35].

This is not quite so impressive if you consider that both Fourier and Taylor expansions both fall into this category, depending on the definition of the arbitrary nonlinear functions $\psi_{ij}(x_i)$. However, Lorentz later showed that this was indeed possible not just with arbitrary functions but also using the *same* nonlinear function for each neuron – *i.e.,* a constant activation function [36]. Thus, in theory, the regression neural network with a single hidden layer of identical nonlinear activation functions should be capable of learning and therefore approximating any function arbitrarily well. Such a network looks something like this:

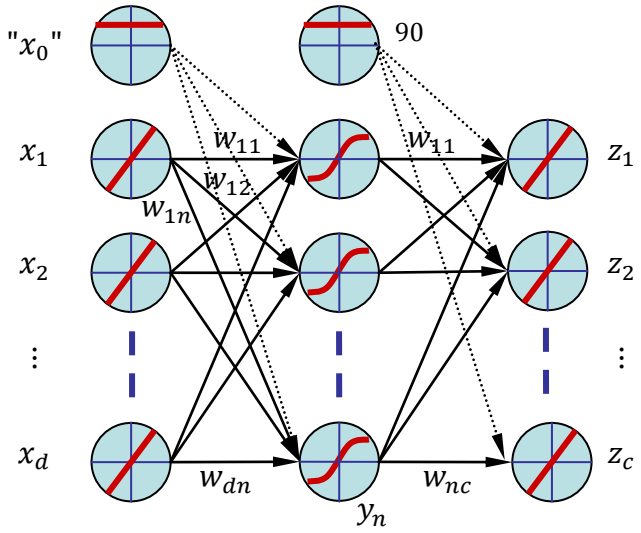



*Figure 3.23: Regression neural network with a single hidden layer. The nodal transfer function is indicated by the graphic. Each arrow carries a numerical weight. Training involves the presentation of known outputs (targets) and known inputs, whereby an algorithm iteratively calculates (backpropagates) the error, updating these weights until convergence on a minimum of the error cost function.*

ANNs are a very flexible and general platform for learning nonlinear functions between almost any conceivable data. ANNs are mentioned in many good texts. Simon Haykin, one of the world's leading experts on adaptive signal processing, gives a comprehensive treatment of neural networks in [37]. Another book by David MacKay [38] gives an interesting background and discusses neural networks with reference to Bayesian learning.

Several years ago a review of neural network techniques applied to microwave measurements (mostly concentrating on calibration for VNAs, but also mentioning materials measurement) stated five advantages of neural networks in this application [39]:

> "(1) they do not require detailed physical models; (2) calibration times can be reduced because only a few training points are required to accurately model standards; (3) ANN model descriptions are much more compact than large measurement files; (4) ANN models, trained

> on only a few measurement points, can be much more accurate than direct calibrations when limited data are available; and (5) they are less susceptible to the noise inherent in the measured data."

Despite being written from the point of view of producing calibrations for VNAs, the same advantages apply to neural networks in materials measurement. The main difference with the sensors is that the training data for the ANN is generated by finite element simulation rather than measurement. This is primarily because enough sufficiently well-characterised materials do not exist, but also has the advantages that the neural network can also be trained to learn the effects of geometry and environmental parameters, like temperature, and thus predict uncertainties in data and to some extend be self-correcting.

Neural network architectures of various types have already been used quite widely in the microwave measurement field, with some applications in permittivity estimation being related to the approach taken in this thesis [40], [41], [42], [43], [44]. Most of the publications on complex permittivity estimation using neural network and other machine learning approaches have been focussed on the problem of arbitrary shaped inclusions in a waveguide under *broadband* excitation. This involves the tricky inverse problem of estimating both the shape and permittivity of an object based on scattering parameter measurements alone. This inverse problem was shown recently, somewhat surprisingly, to have a unique solution [45], allowing it to be used as a primitive form of microwave imaging, although accuracy is surely an issue even with guaranteed singularity. However, reasonable accuracy has been shown for the easier problem of fixed geometry, where a grid of complex permittivity of a static geometry inclusion is taught to a neural network from repeated full 3D finite division time domain (FDTD) simulations of complex reflection coefficient (Figure 3.24). This method was developed by Yakovlev *et al.* [41]. Their neural network fitted the simulated data to an error of less than 0.01% and showed a maximum error of about 5% in actual measurement.

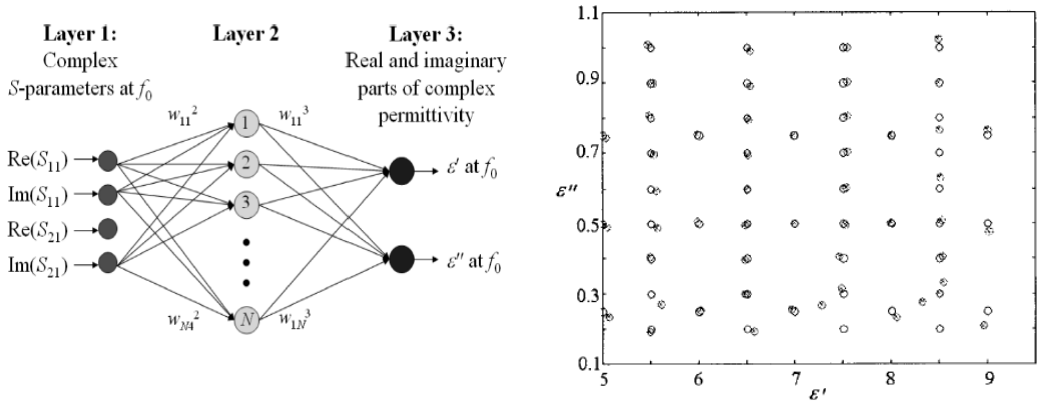



*Figure 3.24: Example neural network structure (left) and sample complex permittivity space (right) showing FDTD simulated training vectors (circles) and corresponding trained network outputs (dots) [41].*

Authors to work specifically on resonator rather than broadband measurements have largely pursued a finite element simulation in the loop approach. In other words, to make a single measurement requires an algorithm to adjust the permittivity of a finite element simulation until it agrees with measured values for the change in resonant frequency and bandwidth, as done in [46], [47]. This is clearly impractical for an online resonant sensor due to the long simulation time, even if neural networks or Newton-Raphson methods are used to speed up the process by locally approximating the finite element results.

Another issue with this, and other neural network methods, is accuracy. Stanković *et al.* [48] recently published a combined knowledge-based network (KBN) and multilayer perceptron (MLP) system for the estimation of the resonant frequency of a tuneable cavity loaded with a dielectric slab. Thus was trained using data generated by the Transverse Resonance Method (TRM), a mode-matching based numerical simulation. They claim a neural network error no better than 0.5 % on average, and a worst-case error no better than about 2 %. This 'model error' would be manifest as distortion in an online MRS, so their technique would need to be improved to be suitable for a precision MRS application. Despite this, their use of a KBN hybrid is notable. First introduced to microwave measurements in 1997 [49] this type of network embeds known analytical or empirical function of the inputs which may describe part of the wider function the network is being taught, making

its task easier. Placing the KBN before a standard neural network effectively generates extra inputs to the network; this may make its life easier, particularly if the knowledge function is highly nonlinear, however there is always the danger that the increased number of weight coefficients needed to accept the additional input dimensions would lead to poor generalisation of the network to unseen patterns.

The apparently simpler problems of estimating permittivity from eigenvalue (the inverse eigenvalue response function, or Stage 2), or estimating eigenvalue from scattering parameters (the inverse resonator spectral response function, or Stage 3) apparently have not been considered independently as candidates for neural network inversion modelling. Yet there are good reasons to divide the task into two like this. Chief among these is that finite element simulation can predict eigenvalues accurately. Another good reason is that, up to a point, eigenvalue simulations are scalable. Thus, the same geometry simulation can be applied to a different frequency, *i.e.,* a slightly larger or smaller resonator of the same geometric ratio (the limit being the error due to incorrect skin effect). In other words, rather than 'tune' the simulation, make the simulation general enough to be applicable over the entire range of anticipated samples and even over multiple geometries if desired.

With sufficient regularisation and control over some finer points of implementation, ANNs and their variations are powerful techniques that can be used to learn complex nonlinear functions, especially where an all-encompassing analytical or empirical function is either non-existent or too computationally intensive to evaluate. However, ANNs extremely slow to train thoroughly (often taking several hours or more with a large dataset) and are only suitable when the underlying nonlinear function is static in time, *i.e.,* training only occurs once. Thus, they are ideal for learning and inverting the nonlinear relationship between, say, resonant frequency, and sample permittivity of a given sensor. In this way, the difficulty of trying to invert a complex nonlinear perturbation equation is sidestepped by training the simulated relationship to a suitable neural architecture, which can then learn the arbitrary inverse relationship.

Stage 3 is the resonator spectral response function, or $\mathbf{S}(\omega, t) = g(\mathbf{p}, t)$, and its inverse is needed for measurement. Curve fitting is traditionally used for this inverse, however neural networks can be applied here too (see Section 5.3.2). This deceptively simple function is deeply connected with the actual method of estimating the spectral response itself. The very nature of this function is dependent not only on the resonator but on the entire readout system. Only by combining neural networks with other powerful techniques from the field of machine learning, such as adaptive filters (Section 5.1.7), can this more challenging equation be accurately represented in silicon.

This concludes this section discussing the theoretical background to resonator modelling. Later sections will draw upon various elements of this theory. The key findings and models discussed will be used to improve various aspects of MRS systems. In Chapters 4 and 5 these are incorporated into the software algorithms and hardware design of a demonstration MRS system which relies on many of the dynamic resonant sensor considerations discussed in this section. In Chapter 6, many of the models introduced here are used to derive important theoretical predictions of MRS response to known permittivity materials. This is a key step in the verification of the newly developed MRS system's accuracy.

## 3.7 Chapter references


[1] D. M. Pozar, *Microwave engineering*. J. Wiley, 2005.

[2] P. K. Day, H. G. LeDuc, B. A. Mazin, A. Vayonakis, and J. Zmuidzinas, "A broadband superconducting detector suitable for use in large arrays," *Nature*, vol. 425, no. 6960, pp. 817-821, 2003.

[3] N. P. L. (Teddington), *Guide to characterisation of dielectric materials at RF microwave frequencies*. National Physical Laboratory, 2003.

[4] Ramamurti Shankar, *Principles of Quantum Mechanics*, 2nd ed. New York: Plenum Press, 1994.

[5] A. S. Eddington, *The Nature of the Physical World*. Kessinger Publishing, 2005.

[6] J. H. Poynting, "On the Transfer of Energy in the Electromagnetic Field," *Philosophical Transactions of the Royal Society of London*, vol. 175, pp. 343 -361, Jan. 1884.

[7] O. Klein, S. Donovan, and G. Grüner, "Microwave cavity perturbation technique: Part I: Principles," *Int J Infrared Millim Waves*, vol. 14, p. 2423, 1993.

[8] R. G. Carter, "Accuracy of microwave cavity perturbation measurements," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 49, no. 5, pp. 918–923, 2001.

[9] R. A. Waldron, "Perturbation theory of resonant cavities," *Proceedings of the IEE - Part C: Monographs*, vol. 107, no. 12, pp. 272-274, Sep. 1960.

[10] J. Venermo and A. Sihvola, "Dielectric polarizability of circular cylinder," *Journal of Electrostatics*, vol. 63, no. 2, pp. 101-117, Feb. 2005.

[11] M. Beleggia, D. Vokoun, and M. De Graef, "Demagnetization factors for cylindrical shells and related shapes," *Journal of Magnetism and Magnetic Materials*, vol. 321, pp. 1306-1315, May 2009.

[12] A. Masood, A. Porch, and D. Barrow, *Microwave Resonators for Highly Sensitive Compositional Analysis*. Lambert Academic Publishing, 2010.

[13] D. Kajfez and P. Guillon, *Dielectric Resonators*. Artech House, 1986.

[14] N. P. L. (Teddington), *Guide to characterisation of dielectric materials at RF microwave frequencies*. National Physical Laboratory, 2003.

[15] A. Porch, *Personal communication*, 2010.

[16] B. Morgan, "Microwave surface impedance of YBCO in the mixed state," 2003.

[17] J. Gao, "The physics of superconducting microwave resonators," 2008. [Online]. Available: http://thesis.library.caltech.edu/2530/. [Accessed: 04-Oct-2011].

[18] C. F. Coombs, *Electronic instrument handbook*. McGraw-Hill, 2000.

[19] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Courier Dover Publications, 1964.

[20] W. B. Case, "The pumping of a swing from the seated position," *American Journal of Physics*, vol. 58, p. 463, 1990.

[21] John C. Butcher and J. C. Butcher, *Numerical methods for ordinary differential equations*. [[John Wiley & Sons]], 2003.

[22] S. S. Haykin, *Adaptive filter theory*. Prentice Hall, 2002.

[23] S. A. Billings, "Identification of nonlinear systems - a survey," *Control Theory and Applications, IEE Proceedings D*, vol. 127, no. 6, pp. 272-285, Nov. 1980.

[24] "Analytic Signals and Hilbert Transform Filters." [Online]. Available: https://ccrma.stanford.edu/~jos/st/Analytic_Signals_Hilbert_Transform.html. [Accessed: 06-Oct-2011].

[25] Hua Zhang, Jianzhong Zhang, Dian Zhou, Jin Liu, Liangjun Jiang, and Yan Pan, "A closed-form phase noise solution for an ideal LC oscillator," in *Proceedings of the 2004 International Symposium on Circuits and Systems, 2004. ISCAS '04*, 2004, vol. 4, pp. IV- 768-71 Vol.4.

[26] P. Langevin, "On the Theory of Brownian Motion," *C. R. Acad. Sci. (Paris)*, vol. 146, pp. 530–533, 1908.

[27] D. S. Lemons and P. Langevin, *An introduction to stochastic processes in physics: containing "On the theory of Brownian motion" by Paul Langevin, translated by Anthony Gythiel*. JHU Press, 2002.

[28] A. Papoulis, *Probability, random variables, and stochastic processes*. McGraw-Hill, 1984.

[29] Gitterman M, "Classical harmonic oscillator with multiplicative noise," *Physica A: Statistical Mechanics and its Applications*, vol. 352, no. 2-4, pp. 309-334, Jul. 2005.

[30] A. P. Gregory and R. N. Clarke, "A review of RF and microwave techniques for dielectric measurements on polar liquids," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 13, no. 4, pp. 727-743, 2006.

[31] H. Cramer, *Mathematical Methods of Statistics*. Princeton University Press, 1958.

[32] R. Penrose, "On best approximate solution of linear matrix equations," *[[Proceedings of the Cambridge Philosophical Society]]*, vol. 52, pp. 17–19, 1956.

[33] B. Fornberg and J. Zuev, "The Runge phenomenon and spatially variable shape parameters in RBF interpolation," *Computers & Mathematics with Applications*, vol. 54, no. 3, pp. 379-398, Aug. 2007.

[34] G. A. Carpenter, M. A. Cohen, S. Grossberg, G. Palm, T. Kohonen, and E. Oja, "Computing with neural networks," *Science*, vol. 235, no. 4793, pp. 1226-1228, 1987.

[35] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley, 2001.

[36] G. Lorentz, "Approximation of functions. 1966," *Rinehart and Winston, New York*.

[37] S. S. Haykin, *Neural networks and learning machines*. Prentice Hall, 2009.

[38] D. J. C. MacKay, *Information theory, inference, and learning algorithms*. Cambridge University Press, 2003.

[39] J. A. Jargon, K. C. Gupta, and D. C. DeGroot, "Applications of artificial neural networks to RF and microwave measurements," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 12, no. 1, pp. 3-24, Jan. 2002.

[40] R. Olmi, G. Pelosi, C. Riminesi, and M. Tedesco, "A neural network approach to real-time dielectric characterization of materials," *Microwave and Optical Technology Letters*, vol. 35, no. 6, pp. 463-465, Dec. 2002.

[41] V. V. Yakovlev, E. K. Murphy, and E. E. Eves, "Neural networks for FDTD-backed permittivity reconstruction," *COMPEL: The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, vol. 24, pp. 291-304, 2005.

[42] H. Acikgoz, Y. Le Bihan, O. Meyer, and L. Pichon, "Neural networks for broad-band evaluation of complex permittivity using a coaxial discontinuity," *The European Physical Journal Applied Physics*, vol. 39, pp. 197-201, May 2007.

[43] M. Maazi, O. Benzaim, D. Glay, and T. Lasri, "Detection and Characterization of Buried Macroscopic Cracks Inside Dielectric Materials by Microwave Techniques and Artificial Neural Networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 12, pp. 2819-2826, Dec. 2008.

[44] H. Acikgoz, Y. Le Bihan, O. Meyer, and L. Pichon, "Microwave characterization of dielectric materials using Bayesian neural networks," *Progress In Electromagnetics Research C*, vol. 3, pp. 169-182, 2008.

[45] Y. Shestopalov and Y. Smirnov, "Existence and uniqueness of a solution to the inverse problem of the complex permittivity reconstruction of a dielectric body in a waveguide," *Inverse Problems*, vol. 26, p. 105002, Oct. 2010.

[46] M. Santra and K. U. Limaye, "Estimation of complex permittivity of arbitrary shape and size dielectric samples using cavity measurement technique at microwave frequencies," *IEEE Transactions on Microwave Theory and Techniques*, vol. 53, no. 2, pp. 718-721, 2005.

[47] K. P. Thakur and W. S. Holmes, "An inverse technique to evaluate permittivity of material in a cavity," *Microwave Theory and Techniques, IEEE Transactions on*, vol. 49, no. 6, pp. 1129–1132, 2001.

[48] Z. Stankovic, B. Milovanovic, N. Doncov, and M. Milijic, "Microwave applicators modeling — Alternative approaches based on neural networks incorporating domain knowledge," in *2010 10th Symposium on Neural Network Applications in Electrical Engineering (NEUREL)*, 2010, pp. 81-88.

[49] F. Wang, "Knowledge-based neural models for microwave design," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, no. 12 PART 2, pp. 2333-2343, 1997.

# 4 Hardware

Both theoretically and in practice, the microwave resonant sensor method can measure a wide range of interesting and relevant dynamic processes accurately and with high precision. Yet a measurement is only as good as the instrument used to make it. Of the numerous existing instruments conceived for characterising, or 'reading-out' resonant sensors, which are most suitable for new and demanding dynamic measurement applications? Apparently, no existing technique is ideally suited; many have a specific advantage, such as cost, speed, or accuracy, but lack the flexibility in architecture to make a smooth trade-off between them.

This chapter discusses the development of a new, more suitable instrument for MRS readout, and, since contemporary systems are as much software as hardware, this chapter focuses on the hardware part of the instrument whilst software is deferred to a chapter of its own. Section 4.1 gives a background to software-defined radio (SDR), the technology behind contemporary communications systems, and the ideal platform for the next generation of microwave resonant sensor systems. Section 4.2 then introduces existing physical implementations of MRS readout systems, discussing how the choice of hardware and its design affects the performance, cost, and integration of the overall system. Section 4.3 then describes a novel design of MRS readout hardware that aims to overcome the limitations of existing systems. We filed a patent for this design and it represents one of the main contributions to knowledge arising out of this work. Section 4.4 then goes on to discuss the experimental implementations of this new readout approach. Completing this chapter are experimental results comparing the static performance of these implementations to a VNA readout method optimised for realtime measurements, along with a short discussion.

## 4.1 Software-defined radio

Fundamentally, all MRS systems rely on the generation (transmission) and detection (reception) of GHz frequency signals. The system passes an excitation through the resonator and detects its response. It then compares the response and excitation in order to characterise the resonator, normally in the frequency domain. When measuring a dynamic environment, the modulation of this excitation signal in amplitude and phase by a time-varying resonator can be seen as a type of analogue quadrature-amplitude modulation (QAM), except with the information being generated by the 'channel' (*i.e.,* resonator) rather than in the transmitter. MRS systems are like high dynamic range radio receivers where the objective is to 'demodulate' this continuous quadrature modulation in order to recover the perturbation signal. Driven by the demands of high-bandwidth communication standards such as Wi-Fi (IEEE 802.11), recent technological advances have greatly increased the sampling rate and analogue performance of QAM-based radios, whilst at the same time achieving a high level of integration, low cost and low power consumption – qualities demanded by a next-generation MRS system. It is therefore logical to look to the latest advances in high bandwidth radio systems and apply them to the problem of MRS readout.

The flexibility, power and, most importantly, the low relative cost of digital signal processing has meant that there is an increasing trend to implement many of the traditional functions of a radio transceiver in digital; minimising the amount of analogue components such as mixers, filters, amplifiers and oscillators. Because "software" in one way or another describes all modern digital systems, radios having a prominent digital element like this are called software-defined radios (SDRs). This trend has been enabled by recent advances in high-resolution and high-speed Analogue to Digital (A/D) and Digital to Analogue (D/A) converters, which now provide high dynamic range (14-16 bit) conversion of 100 MHz signals and up, as well as improvements in the power and speed of digital processors.

Although originally developed in the early '90s by the US Army as a method to integrate many different wireless military communications standards into one flexible device, [1], the term SDR has come to be associated with any radio architecture with a high degree of flexibility, minimalist analogue 'front-end' and a significant amount of digital signal processing. The paradigm of instant reconfigurability to a multitude of existing and unforeseen communications standards is clearly appealing, particularly in military applications where enemy eavesdropping, adaption to variable transmission conditions and communications infrastructures as well as field-downloadable upgrades offer an unquestionable advantage in the theatre of war [2],[3].

In the RF community, it can sometimes seem as if there is an obstructive competition between traditional analogue engineering and digital signal processing approaches to difficult radio challenges. The comparison is tricky, particularly since the old argument that digital systems are low cost and low power tends to be rather weak when dealing with the high performance demands of modern communications. Field-programmable gate array (FPGA) based digital signal processing can easily consume the power budget of an entire receiver system, for example. Therefore, although the SDR paradigm might seem to imply the superiority of a digital implementation over an analogue solution, what it really requires is a careful and cooperative engineering approach that tackles the twin challenges of high performance and power efficiency using the full arsenal of both digital and analogue RF techniques. Only very recently has this convergence of analogue and digital engineering really started to be taken seriously at radio frequencies.

Whilst many of the advantages of SDR systems are irrelevant to MRS systems, a primary benefit from dynamic MRS measurement perspective is the ability to adjust rapidly and accurately the configuration of excitation signal and receiver, eliminating or minimising reliance on a conventional microwave synthesiser or swept sources and their limited tuning rates. The ability for rapid adjustment of RF frequency like this is known as frequency agility, and is a key prerequisite of SDR. Frequency agility facilitates the accurate tracking of rapid changes in resonant

frequency. Digitally generated signals can be updated within one sample period – instantaneously as far as a digital receiver is concerned. As demonstrated in Section 3.8, microwave resonators can respond to changing material properties at rates up to and exceeding their bandwidth, typically several MHz. Therefore, to exploit fully a sensor's potential measurement bandwidth requires the ability to change and detect signals with sub-microsecond timing. High bandwidth digital radio technology is the only practical way to achieve this. Software reconfigurability also allows for dynamic changes in the bandwidth and sampling rate of the detector system, allowing an adaptive trade-off to be made between sampling rate, precision and power consumption. The latter issue is of particular importance for portable and remote sensor systems.

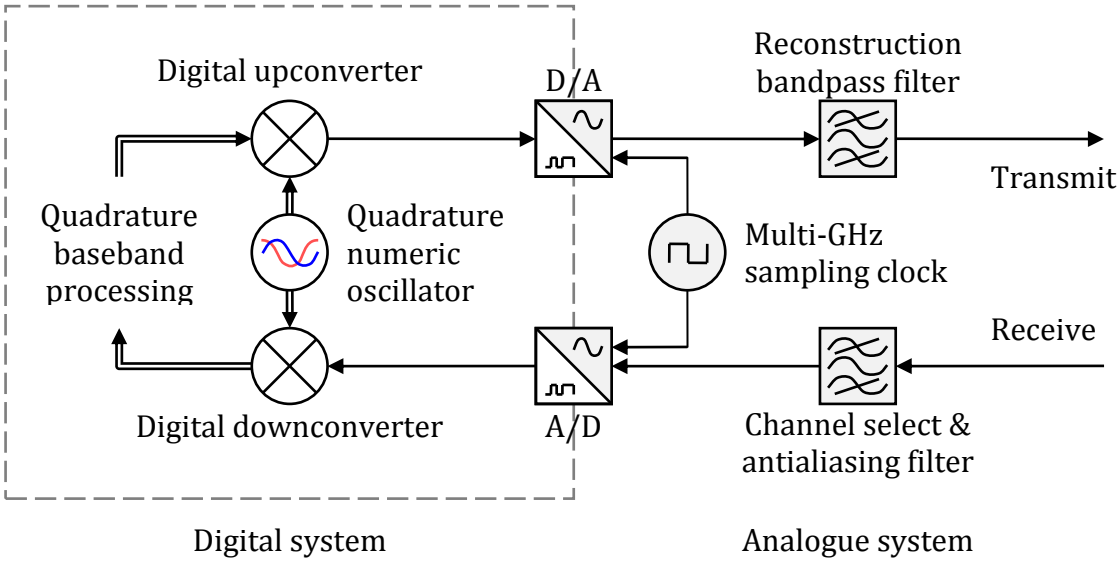



*Figure 4.1: Ideal "full software" radio system with minimal analogue components.*

Despite the advantages of digital systems, it is not currently possible to generate and receive high-dynamic range microwave signals by converting directly to and from digital – the hypothetical Software Radio (SR) (Figure 4.1). Although wideband Arbitrary Waveform Generators (AWGs) exist which perform Digital to Analogue (D/A) conversion at up to 24 GS/s at 10 bit resolution, [4], these instruments are exceptionally expensive (more than £100k) and suffer from poor dynamic range. This is due to the challenging nature of such high frequency

conversion, the imperfections of which generate distortion and wideband spurious signals. On the Analogue to Digital (A/D) side, Digital Storage Oscilloscopes (DSOs) are available for real-time signal acquisition at rates up to 100-120 GS/s and bandwidths as high as 20-45 GHz. However, these can be even more expensive (over £250k) and are usually limited to < 40 dB dynamic range – around 6 effective bits of resolution. The effective number of bits is the equivalent resolution that an ideal converter would need to have to give the same performance; less than the nominal 8 bits due to imperfections such as excess noise, distortion, and spurious signals. Additionally, non-realtime sampling oscilloscopes are available with 80-100 GHz bandwidth [5],[4], and generally have much higher dynamic range than real-time scopes ( ~ 70 dB ). However, they can be quite slow to acquire a single capture as they work by repeatedly (under)sampling a repetitive signal over many cycles. Due to this assumption of static signals, they therefore would not be suitable for the dynamic signals found in MRS systems.

Generating and receiving microwave signals within a finite band is therefore still done most effectively and cheaply using frequency mixing techniques, whereby a lower frequency representation of the desired signal is converted to and from GHz frequencies using an analogue mixer; a nonlinear circuit which generates and filters desired 'mixing products' of the high frequency and baseband signals. In modern radio transmitters, this generally involves using a D/A converter to generate the baseband signal, followed by one or more stages of frequency translation using microwave Local Oscillators (LOs) (Figure 4.2). Receivers work on the same principle in reverse. However, frequency translation with mixers is non-ideal. Spurious signals caused by signal 'images', undesired signal distortion in the mixer, 'folding' of noise into the signal band, conversion losses, LO feed-through, quadrature offset, gain and phase errors and LO phase noise all degrade the performance of analogue front-ends. These issues will need to be understood and addressed carefully so that any detrimental effects can be quantified and minimised in a cost-effective way in order to squeeze the highest performance-cost ratio out of an MRS system.

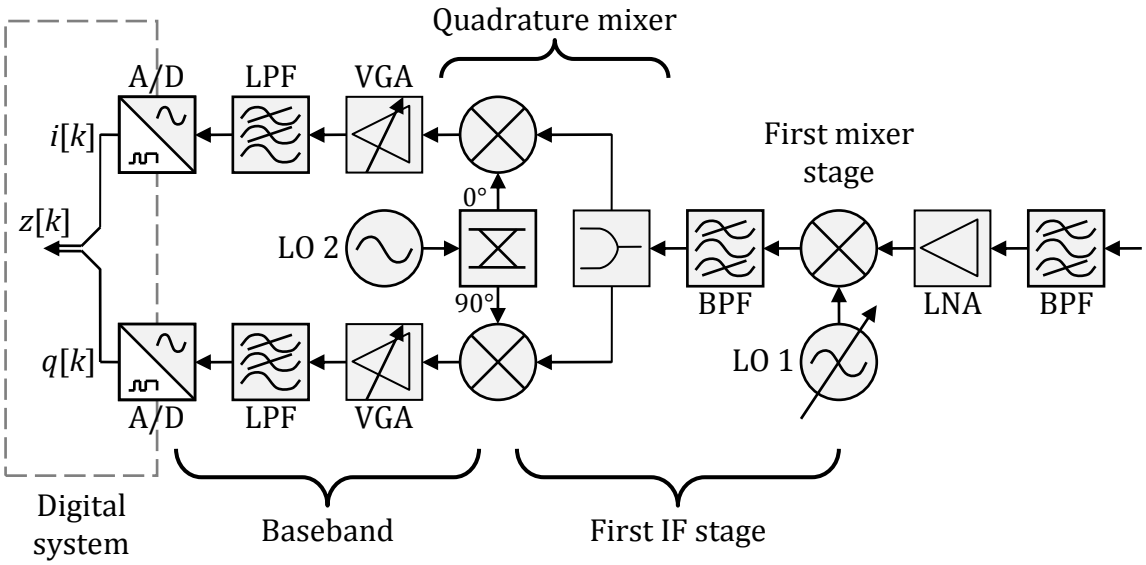



*Figure 4.2: Conventional multistage heterodyne receiver architecture with analogue quadrature downconversion*

The main difference between the requirements of MRS system architecture and radio architectures is the need for analogue measurement accuracy. Unlike a communication system an MRS system is inherently 'closed'; the receiver knows quite precisely, what the transmitter sent and measures the *difference* between what is transmitted and what is received, rather than aiming to recover an unknown transmitted bit stream with a low error. The problem in this differential technique is therefore separating the changes in the received signal caused by the resonator from those due to system imperfections – impedance mismatches, component dispersion, quadrature errors *etc.* Despite this fundamental difference, methods to reduce and correct for system imperfections are still important to improve the error rate of a transceiver system. Thus at a general level, techniques for improving the quality (*e.g.,* data rate, bit error rate) of transmitter systems will also improve the accuracy of MRS systems.

## 4.2 MRS hardware architectures

A variety of different approaches to measuring MRS exists, and each has its own specific advantages and disadvantages. Because the majority of this thesis is concerned with the development of a new MRS readout architecture, it is important first to discuss the most important existing architectures in terms of their operational principles, strengths, and weaknesses. This section will look at these in turn before describing the newly proposed MRS architecture in detail.

### 4.2.1 Power detection

The simplest and earliest MRS architecture combines a swept oscillator with a power detector. The power detector suffers from a number of issues, which arise from its inability to recover the vector (*i.e.,* amplitude and phase) information of signals; therefore, it is largely obsolete except in the most cost-conscious applications. However it is still worthy of discussion as it serves to introduce many of the issues which also affect the performance of more sophisticated MRS systems.

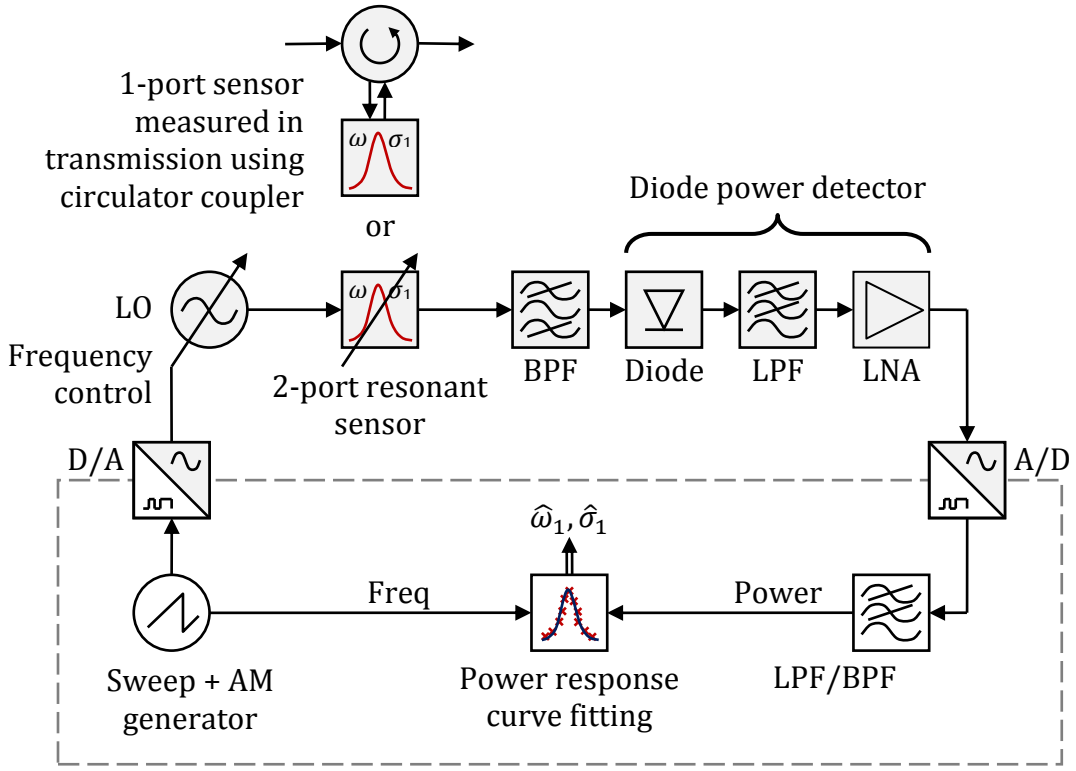



*Figure 4.3: Swept oscillator-power detector architecture*

The main distinguishing feature of this architecture is the power detector; generally a high quality Schottky diode with input impedance matching, DC bias and output filtering, accompanied by a built-in or external 'video' low noise amplifier (LNA). Temperature compensation and response calibration are additional essential features of 'instrumentation' grade power detectors. Instrumentation detectors can have sub-microsecond response time and a typical dynamic range of 60 – 90 dB [6]. They can also be very broadband (*e.g.,* 0.01 – 50 GHz) however, broadband diodes tend to have poor VSWR which leads to load mismatch errors (which distort the resonator response as discussed in the theory section), so narrowband detectors are more suitable for MRS systems. Power detectors range widely in price, but instrumentation quality detectors are relatively inexpensive (typically around £200 – £2000) depending on frequency range and quality (*i.e.,* much cheaper than a VNA). A highly compact detector system using off-the-shelf surface-mount diodes could even be assembled for under £10 if accuracy was not important. However, the total cost of this architecture is usually dominated by the requirement for a rapid sweeping, high

stability and low noise synthesiser – easily costing tens of thousands of pounds. With lower quality, higher phase noise signal sources the precision of this system rapidly deteriorates.

Diode detectors ideally output a voltage proportional to the input power (*i.e.,* using the diode's 'square law' I-V region given a fixed load), which in practice is slightly nonlinear. The general equation for the response of a diode detector is:

$$V_o = I_S R_L \left( \exp\left[ \frac{(V_b - \alpha V_o)}{n V_T} I_0 \left( \frac{A V_i}{n V_T} \right) \right] - 1 \right), \qquad 4.1$$

where $I_S, R_L, \alpha, n$ are constants, $V_T$ is the temperature-dependent thermal voltage, $A$ is the amplitude response of the matching circuit, $V_i$ is the amplitude of the drive signal and $I_0$ is the modified Bessel function of the first kind [7]. This transcendental equation where $V_o$ appears on both sides can be solved iteratively for $V_o/V_i$, giving a response which is approximately quadratic (hence the 'square law') but tends increasingly to a linear relationship at higher power levels. To an extent this nonlinearity can be corrected by calibration. However, there will be some residual systematic error arising from temperature drift of this characteristic, causing distortion of the measured resonator response. To reduce sensitivity to DC drift and close-in LO phase noise, the local oscillator is often amplitude modulated in order to produce an AC signal after the power detector. A traditional analogue lock-in amplifier or digital signal processing then recovers the signal envelope [8]. The response is generally low-pass filtered to reduce noise (at the expense of response time); a digital implementation is preferable as this provides the flexibility to adjust filtering dynamically in response to varying input signal statistics.

The main drawback with this architecture is the non-selective nature of the power detector. Because a power detector integrates the power of all signals, the total noise power over the bandwidth of the detector will be irreversibly added to the desired signal power, as will any spurious signals and harmonics of the signal (hence it is good practice to bandpass filter before the power detector). SNR will

vary over the resonator spectral response if noise is present both on the input and on output of the resonator. In a diode detector, therefore, this variable noise power is added to the signal power, resulting in distortion of the resonator response shape from Lorentzian. In addition, this distortion is not static but changes with sensor perturbation, as the resonator insertion loss and hence SNR changes with changing bandwidth. Thus, this bias is not static and cannot be 'calibrated out', resulting in an unavoidable source of systematic error.

No amount of averaging or lock-in amplification can reduce the bias of a power detection method due to additive input noise. This is an example of bias caused by a nonlinear system with noise on its inputs, a central issue in this thesis. Because of this both the systematic error and the dynamic range of an MRS system based on a diode power detector are significantly worse than is possible with a full vector receiver. For the most part this is a fundamental issue irrespective of the quality (and price) of the components used in either system. The extra cost of precision components tends to go mostly on increased linearity and/or bandwidth.

Bias in power detectors can be demonstrated by simple computer experiment. A sinusoidal signal with added noise was input into simulated vector and diode detectors and averaged 100,000 times (ideally reducing the noise power by 50 dB). This averaging is representative of a very low pass baseband filter or a lock-in amplifier with a very long integration time. Different methods for estimating the power of the input sinusoidal signal were compared. The results confirm that the output of a power detector has bias (shown by its deviation from a straight line), meaning that signals below the input noise floor cannot be resolved. This is exacerbated by the fact that the noise bandwidth of a diode detector must be greater than the maximum anticipated sweep span. However, the output of vector detector with ideal mixers has no such bias, allowing the noise floor to be reduced without limit by averaging – although this is strictly valid only for white noise and ideal system components. This simulation also shows that logarithmic averaging (*i.e.,* the geometric as opposed to the arithmetic mean) is the best way of combining multiple power estimates.

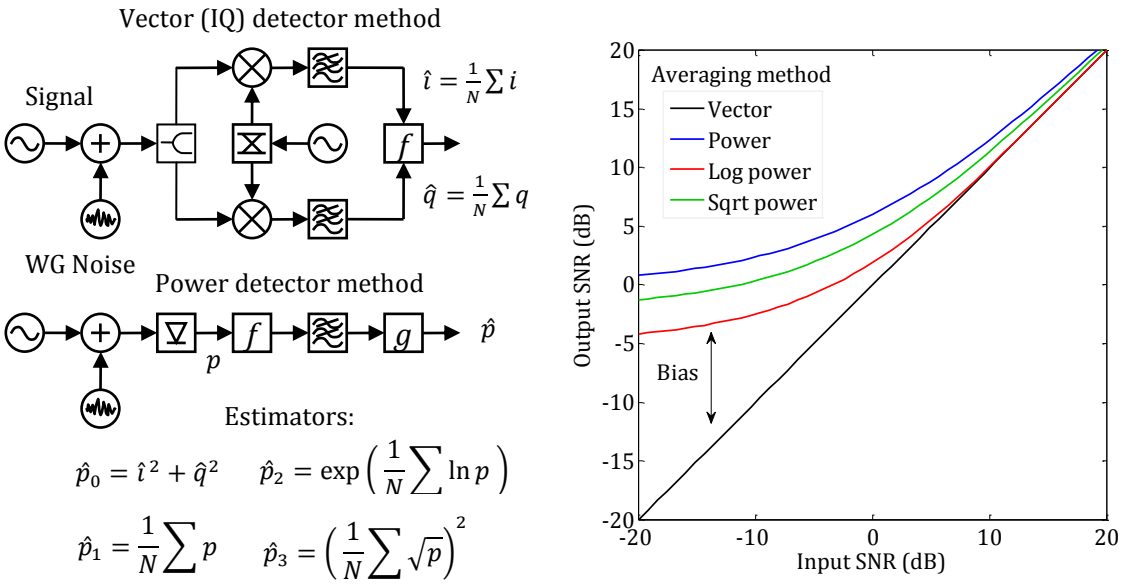



*Figure 4.4: Comparison of different averaging methods for estimating the power of a sinusoidal signal corrupted by white Gaussian noise. Only vector averaging avoids bias and can recover a signal with negative SNR; this requires a tuned receiver. Of the averaging methods suitable for power detector processing, converting the output to a log scale before averaging results in the lowest bias.*

Adding input and output reflectometers to separate the incident and reflected waves turns the system into a 'scalar network analyser' which allows the worst systematic errors to be eliminated through basic calibration routines, though this is in no way comparable to the error correction possible with a vector network analyser. A somewhat better way to keep the low cost and simplicity of diode detectors yet achieve a comparable accuracy to a VNA is to use a six-port reflectometer. The six-port reflectometer technique has been recognised since the 1970s and is reviewed in [9]. More generally, a variety of combinations of 90° hybrids and two, four, or five diode detectors can be used to provide an estimate of the full complex scattering parameter [10]. Since each diode detector of known power but unknown phase gives a circle of possible values on a Smith chart, vector information is found by combining multiple diode detectors offset in phase from each other and finding where the circles they describe intersect. If only one valid solution lies within the Smith chart than this must be the true complex scattering

parameter. More complex circuits and numbers of detectors generally give a more accurate result. Two such reflectometers can therefore provide all four scattering parameters of a two-port network.

The approach has been applied to broadband permittivity estimation on a few occasions [11], [12], particularly at submillimetre wavelengths where mixers are costly and perform poorly [13]. An interesting example of this approach was demonstrated quite recently [14] as a narrowband readout system for a 35 GHz microwave microscope, proposed as a readout for a 'new' method of encoding data as microscopic pits on a disk. Whilst an outdated idea, it was a good demonstration of a microwave microscope nevertheless and was also demonstrated for characterising surface and buried cracks in dielectric materials [15].

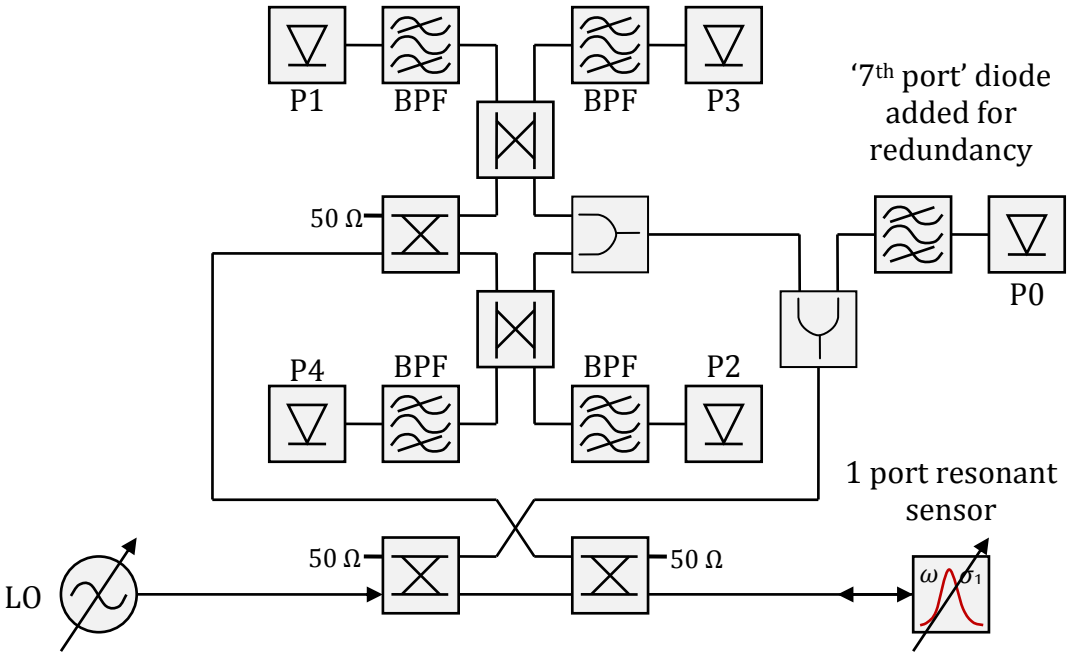



*Figure 4.5: Six port reflectometer using five separate diode detectors, five 3 dB 90° hybrids and two 3 dB 180° hybrids for measuring the complex reflection of an MRS. After [16].*

Despite overcoming the limitations of the receiver in terms of the absolute accuracy (*i.e.,* static error) in scattering parameter estimates, and allowing for the recovery of vector information, the six-port reflectometer method still suffers from

the same drawback as the single diode detector in terms of noise bandwidth. Furthermore, the added cost and size of the extra passive microwave circuits and additional diode detectors can easily become comparable to a full vector receiver. This is especially so if instrumentation quality power detectors are used, although implementing the circuits with lumped element hybrids, *e.g.,* on-chip is a possibility [17]. To overcome the limitations of the diode power detector and provide a better cost-size-performance trade-off, a quadrature receiver can be used instead.

### 4.2.2 Quadrature receiver

Quadrature receivers are behind most modern communications systems, as well as the many of MRS systems, including the new SDR system developed in this work. This architecture also allows for full vector measurement of the response of the resonator. It relies on an In-phase and Quadrature (IQ) mixer, to translate two copies of the received signal down to a lower frequency. Most introductory texts on microwave engineering cover the basics of frequency translation by nonlinear mixing, *e.g.,* [18]. A circuit containing nonlinear components such as diodes, having dominant second order nonlinearity, is used to create a copy of two signals combined at its inputs at both the sum of their relative frequencies, $f_1 + f_2$, and their difference, $f_1 - f_2$. This process is easily demonstrated by considering the series expansion of the voltage input-output relationship of such a device, *e.g.,* $v_o = k_1 v_{in} + k_2 v_{in}^2 + \cdots$. If two sinusoidal voltage waves, an 'RF' signal, $v_1 = V_{RF}\, cos(\omega_1 t)$, and a 'LO' signal, $v_2 = V_{LO}\, cos(\omega_2 t)$, are combined at the input, $v_{in} = v_1 + v_2$, the output voltage can be written:

$$\begin{aligned} v_o = {} & \tfrac{1}{2}k_2(V_{RF}^2 + V_{LO}^2) + k_1(V_{RF}\cos(\omega_1 t) + V_{LO}\cos(\omega_2 t)) \\ & + \tfrac{1}{2}k_2(V_{RF}^2\cos(2\omega_1 t) + V_{LO}^2\cos(2\omega_2 t)) \\ & + k_2 V_{RF} V_{LO}\cos(\omega_2 + \omega_1)t + k_2 V_{RF} V_{LO}\cos(\omega_2 - \omega_1)t, \end{aligned} \tag{4.2}$$

where the last two terms can be seen to be the sum and difference frequencies. The difference frequency thus represents the downconversion process, assuming that $\omega_1 \cong \omega_2$. If instead it is assumed that $\omega_2 \gg \omega_1$ , the same process can be seen to be an upconversion process producing two identical sidebands at $\omega_2 \pm \omega_1$. Thus mixers used for upconversion and downconversion are identical apart from the filtering used to select the desired output signal.

The above analysis considers RF and LO signals with zero phase offset. If instead a signal with a particular phase and magnitude is considered, so that

$$\begin{aligned} v_1 = V_{RF}\sin(\omega_1 t + \phi) &= V_{RF}(\sin\phi\cos\omega_1 t + \cos\phi\sin\omega_1 t) \\ &= V_I\cos\omega_1 t + V_Q\sin\omega_1 t, \end{aligned} \tag{4.3}$$

for an in-phase and quadrature local oscillator (carrier) signal, then the sum and difference mixing products are then

$$\begin{aligned} v_2 = V_{LO}\cos(\omega_2 t): \quad & k_2 V_{LO}\big(V_I\cos(\omega_2-\omega_1)t - V_Q\sin(\omega_2-\omega_1)t\big) + \\ & k_2 V_{LO}\big(V_I\cos(\omega_2+\omega_1)t + V_Q\sin(\omega_2+\omega_1)t\big), \\ v_2 = V_{LO}\sin(\omega_2 t): \quad & k_2 V_{LO}\big(V_I\sin(\omega_2-\omega_1)t + V_Q\cos(\omega_2-\omega_1)t\big) + \\ & k_2 V_{LO}\big(V_I\sin(\omega_2+\omega_1)t - V_Q\cos(\omega_2+\omega_1)t\big). \end{aligned} \tag{4.4}$$

Thus, by using both an in-phase and quadrature carrier signal, and comparing the downconverted products, a receiver can selectively recover the $V_I$ and $V_Q$ parts of a signal. This is evident from looking at the downconverted signals at $\omega_2 - \omega_1 = \omega_{IF}$ in the two baseband channels $i(t)$ and $q(t)$:

$$\begin{bmatrix} i(t) \\ q(t) \end{bmatrix} = k_2 V_{LO}\begin{bmatrix} V_I\cos\ \omega_{IF}t - V_Q\sin\omega_{IF}t \\ V_I\sin\ \omega_{IF}t + V_Q\cos\omega_{IF}t \end{bmatrix} = k_2 V_{LO}\begin{bmatrix} V_I & -V_Q \\ V_Q & V_I \end{bmatrix}\begin{bmatrix} \cos\ \omega_{IF}t \\ \sin\ \omega_{IF}t \end{bmatrix}, \tag{4.5}$$

which is a special matrix multiplication which, as may be recalled from Section 3.3, is equivalent to complex multiplication, $i(t) + jq(t) = \big(V_I + jV_Q\big)e^{j\omega_{IF}t}$, where the complex constant $\big(V_I + jV_Q\big)$ is the Fourier coefficient of the RF signal. The Fourier components of the RF signal can thus be recovered irrespective of the relative

phases of the RF and LO waveforms. This is crucial because phase synchronicity of microwave signals is very difficult, if not impossible, to achieve over reasonable time scales. Therefore, phase and magnitude spectral response information of a resonator can be recovered simply by sweeping RF frequency.

Although the six-port reflectometer also enabled the estimation of the full complex scattering parameter, the homodyne quadrature receiver has one big advantage over diode detectors. Because the frequency mixing process ideally translates a linear copy of the signal with minimal distortion, the noise can be reduced by filtering or averaging at baseband, effectively allowing a much smaller noise bandwidth than possible with diode detector. This is only possible with a linear detector such as this. Spectrum analysers utilising this coherent detection principle are often specified with noise bandwidths as low as 1 Hz and consequently can have as much as 150 dB of dynamic range. Without doubt, this is a significant advantage, and arguably does not involve additional complexity. After all, a quadrature, or IQ, mixer generally requires 4-8 diodes to form two individual mixers, a 90 hybrid to create the in-phase and quadrature carrier signals, splitters to send the RF signal to each mixer, and low pass filters to remove the unwanted higher frequency products. Clearly this is a similar level of complexity to the six port diode reflectometer considered earlier, and is based on largely the same components having the same performance-cost considerations.

If the frequencies of these two signals are the same, the process creates a DC voltage and a signal at $2f$. This is known as homodyne conversion, after the Greek words *ὅμοιος* (*homoios*) meaning similar and *δύναμις* (*dynamis*) meaning power or force; the root of the word "dynamic". The term 'zero-IF' is also used for this configuration since the intermediate frequency (IF), defined as $\omega_2 - \omega_1$, is zero and no further frequency mixing stages are used. In the equation above, if $\omega_2 = \omega_1$ the DC components reduce to $k_2 V_{LO} V_I$ with the in-phase carrier and $k_2 V_{LO} V_Q$ with the quadrature carrier. Using a hybrid to generate these 0° and 90° carrier signals, it is therefore possible to recover the relative amplitudes of $V_I$ and $V_Q$ and therefore

the phase shift of the RF signal relative to the carrier. This allows the full complex resonator response to be estimated.

However, referring to the series expansion of the nonlinear mixing process above, an issue with this 'DC' homodyne approach becomes apparent. The nonlinearity of the mixer itself produces the undesired DC components, $\frac{1}{2}k_2(V_{RF}^2 + V_{LO}^2)$. These will be superimposed on the desirable homodyne DC levels giving the real and imaginary components of the signal, leading to an error. In fact the total DC levels on the $I$ channel will be $k_2\left(V_I^2 + V_Q^2 + V_{LO}^2 + 2V_{LO}V_I\right)/2$ and $k_2\left(V_I^2 + V_Q^2 + V_{LO}^2 + 2V_{LO}V_Q\right)/2$ on the $Q$ channel. Furthermore, since the LO drive level can be 10 dBm or more, especially for passive mixers, the DC output due to this term would swamp that due to the desired signal. This not only leads to errors due to drift, and noise from imperfect LO automatic gain control, it also creates the potential problem of saturating the input amplifiers and/or ADC. To get around this problem, balanced mixers are required. A typical balanced mixer consists of a diode ring and two 180° hybrids or centre-tapped transformer baluns:

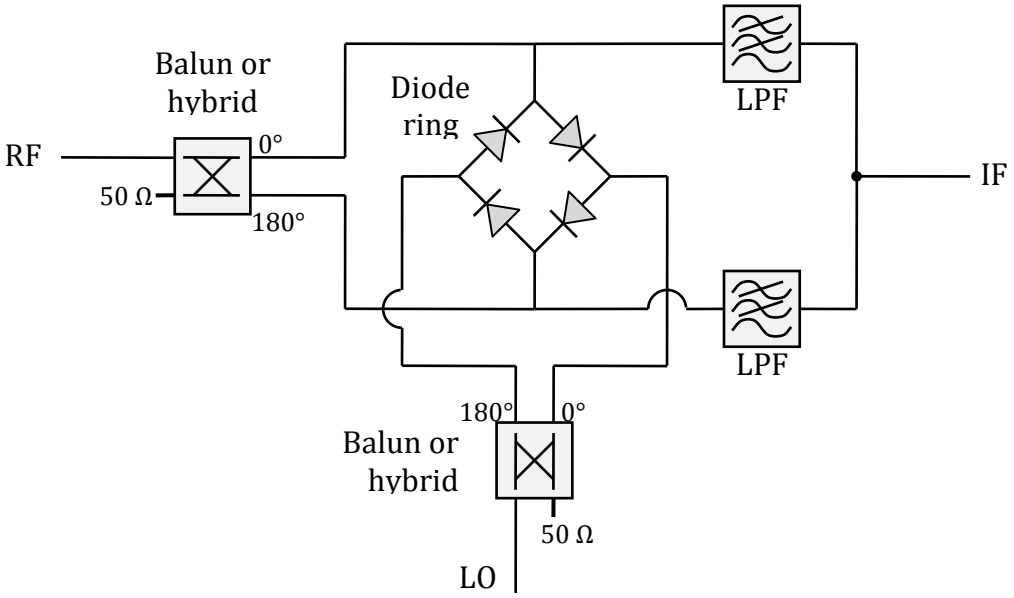



*Figure 4.6: Double balanced mixer circuit based on a diode ring and 180° hybrid transformers on both the RF and LO ports.*

The balanced mixer ideally cancels the undesired DC components, and in fact all even order harmonics of the RF and LO signals [18]. This compromises input impedance matching on the RF, however. Circulators or directional couplers are necessary to prevent resonator response errors due to load mismatch. Furthermore, it is impossible to design ideal 180° hybrids, especially as the diodes themselves may be mismatched due to process variations. Thus whilst this circuit does alleviate the problem, it does not eliminate it. Mixers are usually specified for LO to IF isolation. The IQ mixers used in the prototype system have 30 dB LO-IF isolation, so for an LO drive level of 10 dBm, RF signal levels any smaller than -20 dBm will be swamped by LO feedthrough. Thus, this mixer specification must be excellent for the homodyne approach to be accurate and free from the potential for drift errors. Diode temperature variations can cause time-varying drift, affecting this match and therefore the level of undesired DC feedthrough. This could be confused with a small real-world signal, especially if single frequency readout is used.

To avoid this problem altogether requires modulated signals in place of just the LO itself as the excitation, or a full heterodyne receiver architecture having two or

more local oscillators and several frequency translation stages. In both these cases, a bandpass filter at IF filters out DC. These methods are used in the SDR architecture and in VNAs, respectively.

Another significant source of error in quadrature receivers is gain and phase imbalance, or quadrature skew, between the I and Q channels. This error also affects the SDR system, so will be considered in more detail later on. Due to imperfections in the quadrature hybrid, and gain differences in the mixers and subsequent baseband processing, the true $i(t)$ and $q(t)$ signals received will be erroneous. Instead of representing exactly orthogonal phasors of equal magnitude, there will be a small error in their relative magnitudes and the angle between them. At a given frequency, this error is represented by linear transformation of the $i(t)$ and $q(t)$ signals:

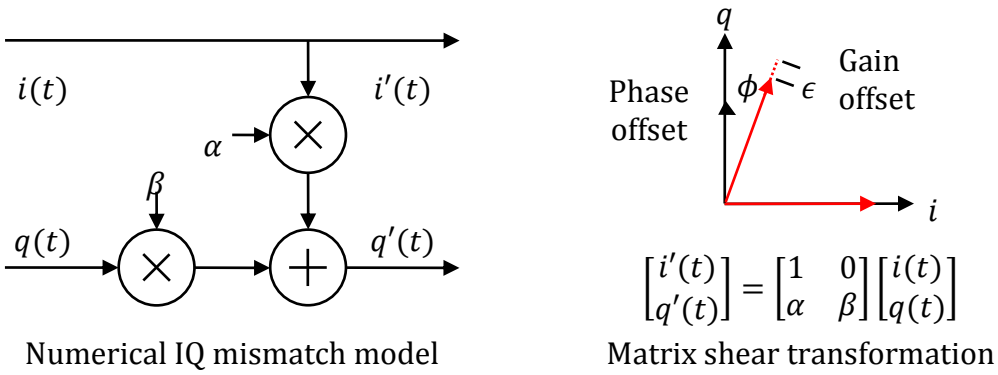



*Figure 4.7: Representation of IQ mismatch error; digital circuit representation (left), phasor (upper right) and matrix multiplication, or shear transform, representations (lower right) adapted from [19].*

This error means that the complex number equivalence is broken, as the matrix representation no longer has the correct form (see Section 3.2). In homodyne systems, this leads to systematic error in spectral estimation, particularly as this error varies with frequency. However, as can be discerned from the matrix representation, this error is linear and the true $i(t)$ and $q(t)$ signals can be recovered by multiplying by the inverse of the shear transform matrix. An operator can estimate the coefficients of this correction matrix by comparison to a known

reference – for example by connecting a 'thru' line in place of the resonator with a known phase delay. Whist this is not easy to do at DC in a true homodyne system, it is possible to use a variable length transmission line and sweep the phase of either the resonator or the reference branch by changing the line length. This should ideally trace out a circle in the IQ plane. A calibration routine can then estimate the quadrature imbalance coefficients from the measured deviation from this circle. Clearly, these calibration requirements are unsuitable for MRS systems outside of a laboratory.

Simple blind estimation methods such as those described in [19],[20],[21] exist that do not require a calibration to a known reference and work under the statistical assumptions that the true $i(t)$ and $q(t)$ signals are random and uncorrelated. Whilst acceptable for communications systems, such algorithms do not provide the degree of accuracy required for resonator measurements and the time-domain variation caused by the stochastic nature of their adaptation could be confused with a real perturbation signal. On the other hand, higher performance blind correction methods such the Bussgang or Constant Modulus algorithms [22] are aimed at digital communication systems employing a discrete symbol alphabet and are unsuitable for the unknown analogue perturbation encountered in this system. Using calibration to a known reference, however, leaves the possibility of errors due to drift, and it is generally not practical to recalibrate during a dynamic measurement when it requires the physical disconnection of the sensor.

Figure 4.8 below shows a design for a simple MRS readout system based on a simple homodyne quadrature receiver. The LO excitation is generated by a synthesiser which should have good amplitude stability and low phase noise. It is filtered to eliminate harmonics and then split using a directional coupler into a resonator branch and a reference branch which provides the LO drive signal for the mixers. The coupler is preferable to a simple resistive divider as it ensures good isolation of the LO reference arm from reflections due to the resonator. It also provides a more appropriate distribution of LO power since the mixer (if passive) typically requires a high drive level which is undesirable for the sensor due to self-heating effects. The diagram shows separate balanced mixers configured as an IQ

mixer, but an integrated passive or active device could be used instead. LNAs, preferably with variable gain, are used to provide an optimum signal level to the two channel A/D converter. The low pass filters provide both RF signal rejection and antialiasing and must therefore have good rejection over a wide frequency range. Finally the D/A converter shown allows for frequency control of the LO through by some unspecified method, allowing the digital processing to control the sweep in a synchronous way for resonator spectral characterisation. Swept characterisation is essential when using the homodyne architecture for calibration purposes, even if single frequency readout is used subsequently for speed. This architecture and calibration method was the approach taken with the first kinetic inductance detector architectures, for example [23].

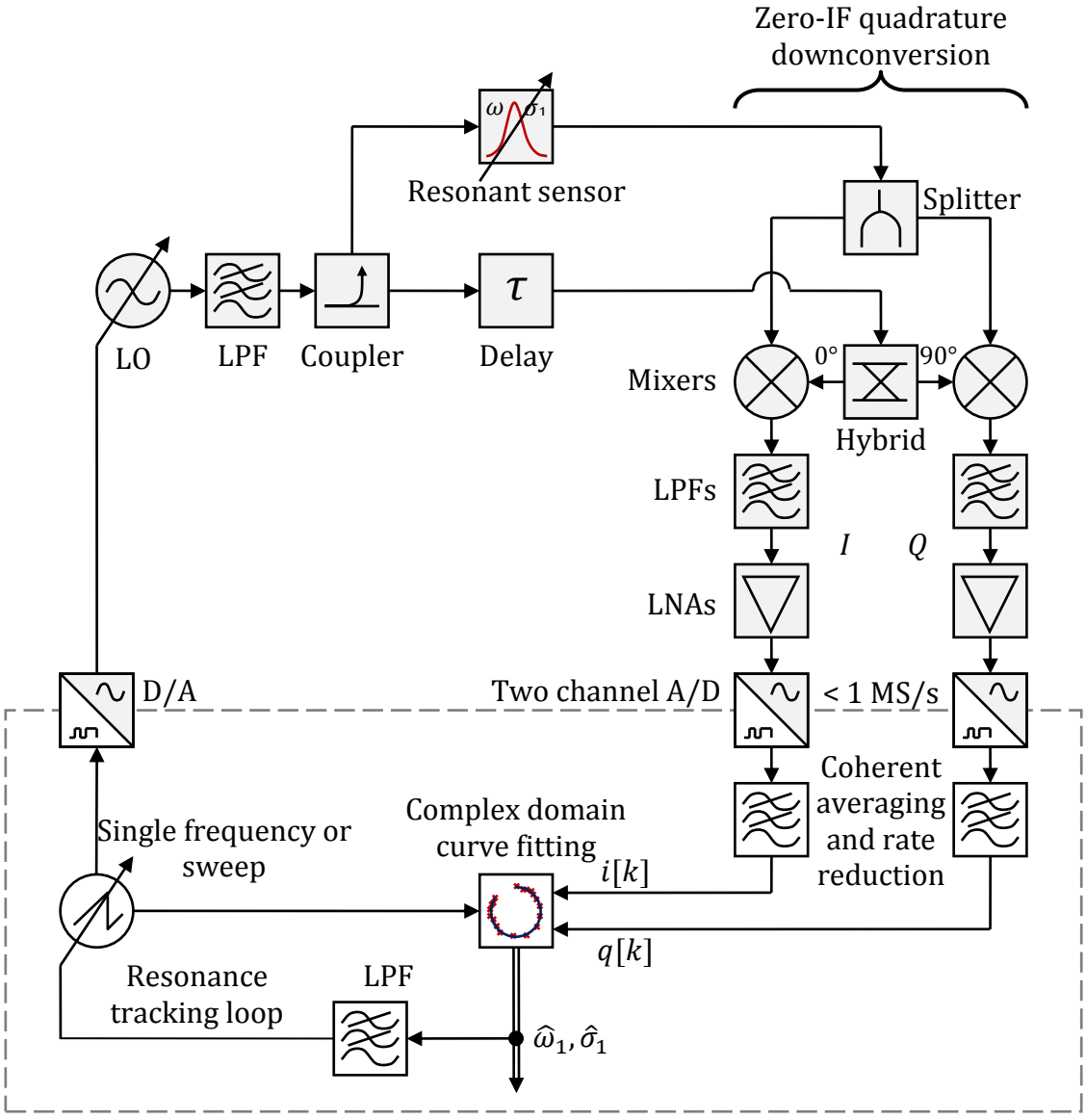



*Figure 4.8: Homodyne quadrature detector based microwave resonant sensor readout architecture.*

The A/D converters here are nominally sampling signals with a bandwidth equal to the highest frequency content of the perturbation signals expected for the resonant sensor. Because this is usually much less than the resonator bandwidth, it makes sense to use a lower rate, higher resolution converter in this architecture. An ideal A/D converter for this is Σ-Δ. This architecture uses a faster A/D to oversample the analogue waveform combined with noise shaping to push the quantisation noise away from the lower frequency band of interest. Modern, multilevel Σ-Δ ADCs can achieve 24-bit resolution at rates as high as 1 MS/s – easily fast enough for most

dynamic MRS needs. An example of such a device is the AD7760 from Analogue Devices [24]. This device has over 100 dB dynamic range at 2.5 MHz (8 times oversampling). The oversampling factor can also be programmed from 8 – 256, providing the system with a way of trading off SNR against maximum input frequency on the fly. This component would be an ideal choice for high performance dynamic MRS measurement applications using the homodyne architecture, although with a power consumption of almost 1W it is not suitable for portable applications.

Note that a compensating delay can be included to ensure that the phase of the signals in the LO is matched to that of the resonator branch. Whilst phase is relative and any phase offset can be corrected via calibration, it is tempting to think that having equal delays in these branches should cancel out any phase noise generated by the LO. In theory at least, this random phase modulation superimposed on the LO signal can be cancelled by the homodyne method if the electrical delay, or group delay, in both the resonator branch and the reference branch are equal. One way to think about this is to consider a time-varying 'IF' frequency $\omega_2(t) - \omega_1(t)$ where $\omega_2(t) \approx \omega_1(t)$ and both $\omega_2(t)$ and $\omega_1(t)$ are subject to random variations of frequency. This is an approximate model for CW signals derived from the same source but offset slightly in time (frequency is the derivative of phase; so a phase noise distribution of $-20$ dB/dec would translate to white frequency noise). The larger any delay offset, the higher the amount of LO phase noise present at the receiver. Quadrature homodyne mixing with a phase offset like this is even used as a simple way to measure oscillator phase and amplitude noise without needing a reference oscillator [25].

This suppression effect can be shown with a simple simulation where a small phase modulation is added to a local oscillator signal, $v_2 = V_{LO}\cos(\omega_2 t + \epsilon\sin(\omega_m t))$, and a variable delay is introduced into one branch. Ideal quadrature mixing is simulated mathematically: $z(t) = v_2(t-\tau)\cdot\mathcal{H}(v_2(t))$, where $\mathcal{H}(.)$ denotes the Hilbert transform. From a real input signal, the Hilbert transform produces a complex signal where the imaginary part is equal to the input but shifted by 90°. After putting the simulated carrier through this mathematical

hybrid, the resulting multiplication with its delayed copy, $z(t)$, then has a real part equivalent to $i(t)$ and imaginary part equivalent to $q(t)$.

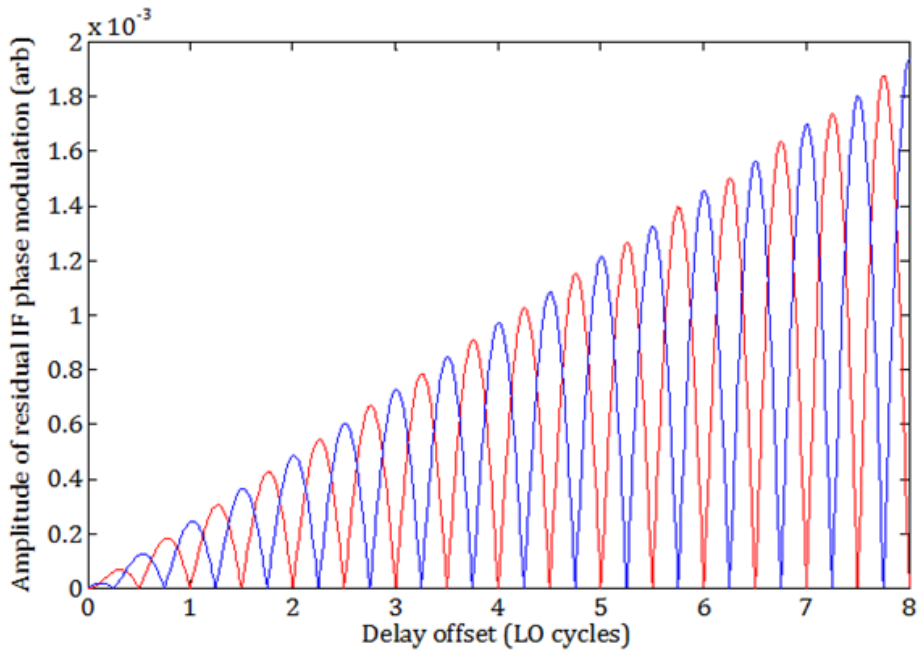



*Figure 4.9: Simple demonstration of phase noise cancellation in a homodyne detection system for increasing delay offset of the two branches. Red shows the amplitude of the residual phase signal in the I channel, blue the signal in the Q channel.*

The results confirm that as the delay is increased the suppression of phase modulation decreases. The residual phase noise present on the I and Q channels can be seen to follow the relationships; $i \propto \omega_c \tau \, sin^2(\omega_c \tau)$, and $q \propto \omega_c \tau \, cos^2(\omega_c \tau)$. Therefore, the overall phase noise magnitude on both channels combined is simply proportional to the delay, $\tau$. Whilst this does suggest that minimising $\tau$ is important for phase noise rejection, this is almost impossible to achieve practically when a resonant sensor is in place because its group delay is very large. Group delay is defined as the derivative of phase with respect to frequency:

$$\tau_g(\omega) \triangleq -\frac{d\phi(\omega)}{d\omega}, \quad \phi(\omega) = \arctan\frac{2Q_1(\omega_1 - \omega)}{\omega_1} \implies \tau_g(\omega_1) = \frac{2Q_1}{\omega_1}, \qquad 4.6$$

the resonator offset constant, $K_0$, has assumed to be zero for simplicity. Thus, the group delay of a resonator measured in RF cycles is about $2Q_1$, over 100 ns for most practical resonant sensors of interest, the same as about 21m of PTFE-filled coaxial cable. This delay also changes with sample perturbation. This all implies that it is impractical to cancel phase noise by delay compensation, and suggests that resonator measurements are inherently sensitive to LO phase noise due to their large group delay. This inherent sensitivity to LO phase is one of the main reasons why it standard practice to use the lowest phase noise oscillator affordable in a given MRS application.

When the LO frequency is swept, accurate estimates for the resonator gain and magnitude response are possible, as long as the receiver is calibrated to correct for IQ gain, phase and offset errors. Figure 4.8 above shows complex-domain curve fitting in the digital signal processing section as a means to estimate resonator parameters. Complex-domain fitting makes full use of the complex response information available. Yet even if the software converts spectral data to power and uses Lorentzian fitting instead, it does not mean that the IQ system is a 'waste'. The enhanced accuracy and lower noise bandwidth still give this system architecture a significant performance advantage over power detection methods.

Swept frequency characterisation is necessarily slow, as the LO synthesiser must make a difficult trade-off between sweep speed, phase noise and resonator settling time. To improve the readout rate for realtime measurements, the system can use single frequency (*i.e.,* continuous wave, or CW) excitation instead. The signal processing can make use of Equation 3.81 in this case:

$$\omega_1(t) = \frac{2\big(\omega - \omega_1(0)\big)\tilde{p}(t) + \omega_1\tilde{q}(t)}{2\tilde{p}(t) - \tilde{q}(t)}, \qquad \tilde{p}(t) = \tilde{\imath}^2(t) + \tilde{q}^2(t)$$

$$\sigma_1(t) = \frac{\omega\tilde{\imath}(t) + \sigma_1(0)\big(\tilde{q}(t) - 2\tilde{p}(t)\big)}{2\tilde{p}(t) - \tilde{q}(t)}, \qquad \tilde{\imath}(t) = (i(t) - K_0)/K_1\,, \quad \tilde{q}(t) = (q(t) - K_0)/K_1 \tag{4.7}$$

This relates both resonant frequency and bandwidth directly to the time-domain $i(t)$ and $q(t)$ waveforms, allowing the system to make resonant frequency and bandwidth estimates up to the full ADC sample rate. However, this requires a calibration routine to correct the waveforms for quadrature imbalance errors, normalised by the measured resonator $K_0$ and $K_1$ constants and referenced to the static resonator frequency $\omega_1(0)$ and bandwidth $\sigma_1(0)$ (estimated from a previous swept-frequency measurement or VNA characterisation). It is also suitable only for small changes in resonant frequency and bandwidth, otherwise accuracy quickly deteriorates.

Whilst the quadrature homodyne system makes a number of improvements over power detection methods, it evidently still suffers from a number of issues. The ultimate and most accurate way of measuring a linear microwave network is the vector network analyser. Correcting both for system imperfections such as source and load impedance mismatch, as well as using a heterodyne receiver architecture for both vector information recovery and high dynamic range, it is undoubtedly the most accurate and most precise way of characterising a linear network such as resonator. However, this comes at a price. The next section looks at the VNA and its architecture

### 4.2.3 The vector network analyser (VNA)

Rather than use a homodyne quadrature receiver, modern VNAs use a CW excitation source offset in frequency from the LO of the receiver, creating a MHz frequency IF signal that can be bandpass filtered with a high rejection crystal filter. Careful and adaptive choice of this IF frequency minimises the number and magnitude of spurious signals that fall into the measurement band, resulting in a very high dynamic range. In modern VNAs this IF signal is then generally converted directly to digital, before being processed using digital downconversion and filtered according to the user-specified IF bandwidth (IFBW). The digital downconversion process implements a numerical equivalent of a homodyne quadrature receiver, utilising efficient frequency domain processing such as hardware FFTs on a dedicated FPGA chip.

A heterodyne downconversion system therefore allows for full recovery of the phase and magnitude information of a signal without requiring an IQ mixer. In fact some VNAs, such the now obsolete Agilent 8753 series [26] did away with a mixer altogether and used an RF sampler instead, which under-samples the RF signal and converts it to baseband through a wideband aliasing process. Conceptually this is similar to mixing by an impulse train, or a 'comb spectrum' as this would appear in the frequency domain. Apparently, this proved to be cheaper despite the extra complexity, probably because it is difficult to make the very wide bandwidth mixers required of VNA receivers:

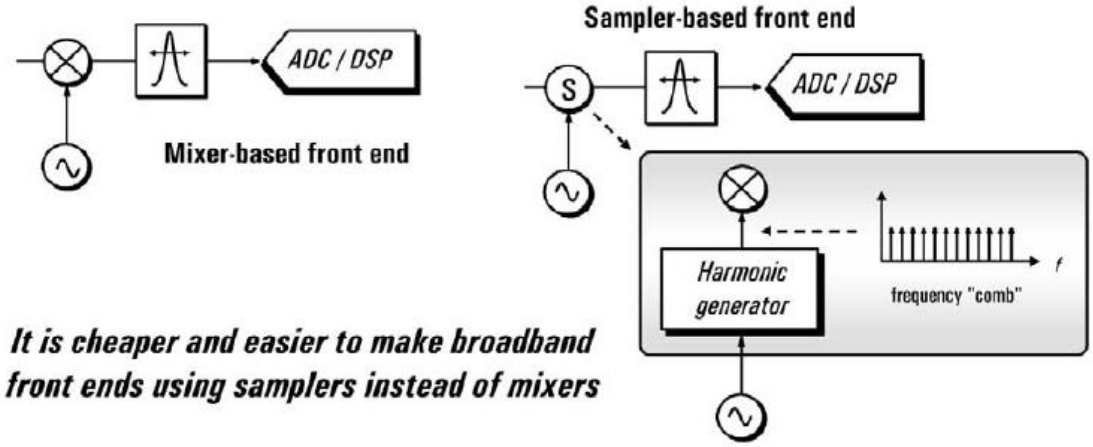



*Figure 4.10: VNA receiver 'front-ends' are based on wideband samplers rather than a traditional mixers in some economy Agilent VNAs despite the degradation in dynamic range due to noise folding in from over its wide bandwidth [26].*

At its heart, the VNA consists of a minimum of three coherent, heterodyne receivers. Directional couplers on the outgoing and incoming signal paths allow it to measure the three voltage waves in a given direction, $a_1$ – the excitation wave, $b_1$ – the reflected wave and $b_2$ – the transmitted wave. By forming vector ratios it can determine the scattering parameter estimates $\hat{S}_{11} = b_1/a_1$ and $\hat{S}_{21} = b_2/a_1$. The LO excitation and 50Ω load are then swapped with a switch and the two reverse scattering parameters calculated. This can be expanded indefinitely and four port VNAs are common. The receiver architecture and overview of a typical VNA is shown in the Figure 4.11.

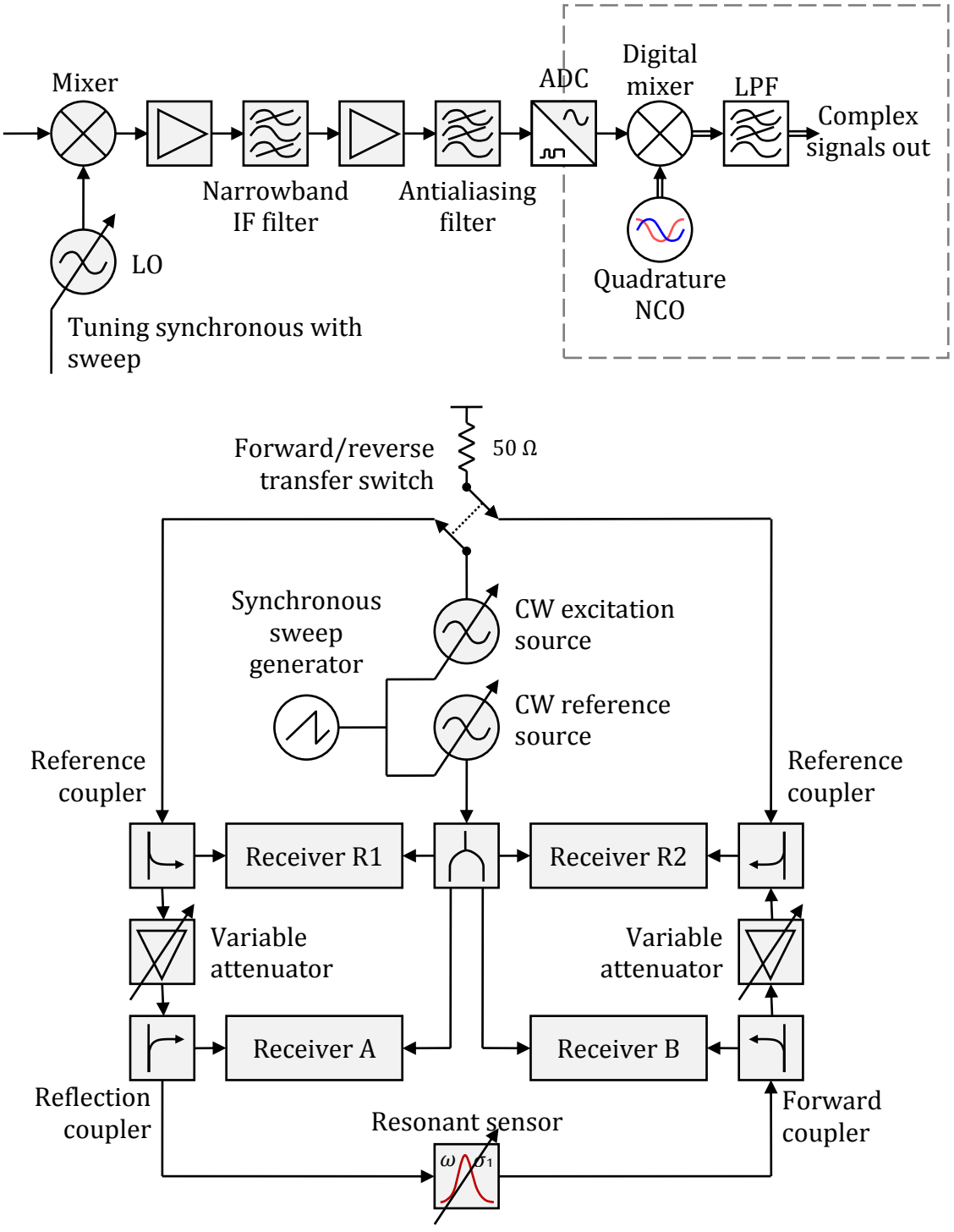



*Figure 4.11: Typical receiver architecture of a modern VNA (top), and a 2-port VNA architecture based on four receivers and a single excitation source used for MRS readout (bottom), adapted from [27].*

Even without calibration, the use of the vector ratio method and the high quality of the internal couplers and components mean that the base accuracy of the VNA is reasonably good, as demonstrated by the following figures for the economy model VNA used for some of the later experimental work. Evidently, the only rather poor specification is the load match on this particular instrument.

*Table 4.1: Specifications for sources of measurement error before calibration for the Agilent E5071B used for the experimental work [28].*

| Parameter | Specification (dB) | | |
|---|---|---|---|
| | 0.0003 – 3 GHz | 3 – 6 GHz | 6 – 8.5 GHz |
| Directivity | 25 | 20 | 15 |
| Source match | 25 | 20 | 15 |
| Reflection tracking | ± 1.0 | ± 1.5 | ± 1.5 |
| Transmission tracking | ± 1.0 | ± 1.5 | ± 1.5 |
| Load match | 15 | 11 | 8 |

Due to the number of non-ideal components in the signal path – couplers, switches and the load impedance itself, a 6 complex term (*i.e.,* 12 real term) error model is used to calibrate VNAs using known reference standards, such as a short, open, load and through line normally used in a coaxial environment. This error correction provides a greatly enhanced 'virtual' directivity, source match, *etc.*, typically better than 40 dB for directivity and match and reducing tracking errors to ± 0.05 dB or better [6].

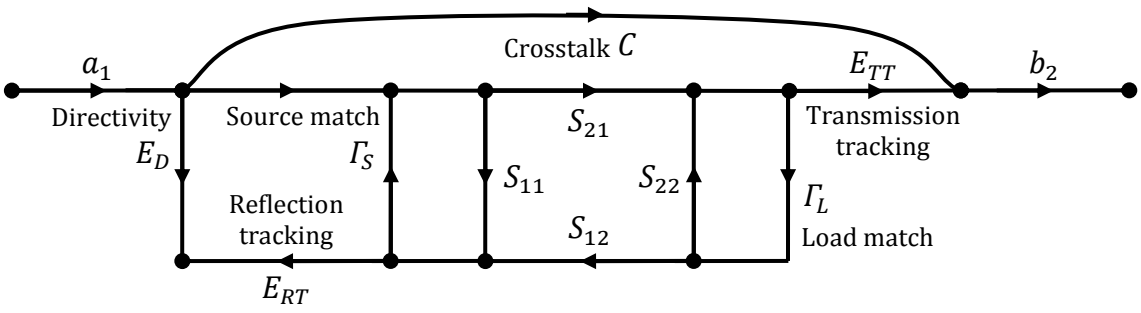



*Figure 4.12: Error flow graph for VNA forward measurements*

However, these errors are not necessarily easy to translate to resonant frequency or bandwidth estimation error in MRS measurement. Many of these sources of error are frequency dependent but vary relatively little over the narrow bandwidth of a typical resonance, even over the entire perturbation range of a particular sensor. As discussed in the theory section, if the errors are constant to first order with frequency it causes a quadratic distortion of the resonator response, leading to bias in bandwidth estimates. Yet in a differential resonator measurement, some of these errors may cancel out. Simulation studies were carried out to attempt to address this issue, as discussed in the Software chapter. It is likely that VNA calibration does not have a significant impact on MRS accuracy, especially when compared to other sources of error such as resonator drift and geometric tolerances. This renders much of the complexity of the VNA unnecessary. In particular, reverse measurements are not required to make transmission and reflection measurements when calibration is not being used. This effectively wastes two of the four receivers.

Whilst the architecture of a VNA is not overly complex in concept, it is expensive because of the duplication of the receivers and the need for two precision RF sources. Because VNAs are designed to be very broadband and have very high specifications throughout this band, the individual RF components such as couplers and mixers (if used) are expensive and bulky. For MRS systems, which for a given application or sensor do not require broadband operation at all, this makes the VNA an over specified and non-cost effective solution.

## 4.3 Swept oscillator considerations

From the earlier discussions, it is clear that swept oscillator measurements are central to all exiting MRS readout systems. The frequency of the source must be swept in a power detection mode otherwise changes in bandwidth and frequency cannot be resolved independently, even in the small perturbation limit. It is also necessary to sweep the LO in homodyne quadrature readout systems for calibration, and repeat this periodically to correct for drift errors. Similarly, although single frequency, or continuous wave (CW), readout is possible in VNAs that support it, this also requires periodic swept-measurements because of the difficulty tracking changes in resonant frequency with a CW excitation. Periodic calibration is unsuitable for continuous real-time measurements. The new MRS system designed and built for this thesis is capable of spectral characterisation without requiring a swept oscillator. With existing technology, there are probably four options for dealing with swept oscillator characterisation in the context of dynamic measurements:

- Sweep the oscillator every measurement cycle and live with limited readout rate (meaning longer scan times for microwave microscopes, inability to capture important dynamic behaviour, reduced precision, *etc.*);
- Keep the CW frequency on resonance using a tracking loop, sweeping the oscillator only to establish an initial lock or in the event of a lock failure;
- Use a fixed frequency CW excitation and control the sample and environment conditions carefully to ensure resonance is always within a narrow measurement window ;
- Sweep the oscillator periodically for calibration, using a fixed frequency CW excitation for discontinuous yet higher-rate measurements.

The first option raises a fundamental issue in all swept readout systems. This is because sweep speed, frequency accuracy and phase noise are inextricably correlated to the performance (*i.e.,* expense) of the LO. The PLL bandwidth should be low to ensure good phase noise but high to allow the oscillator to settle quickly to a new frequency when swept or stepped. Only by using an inherently lower

noise VCO or higher performance PLL architecture can this trade-off be improved. This makes the LO the limiting component in the overall figure of merit of the MRS system. Because of this, subsequent improvements to the aforementioned readout architectures have generally involved the use of better types of LO, such as those based on direct digital synthesis (DDS). This indicates that there is a big advantage in keeping the LO frequency fixed in a particular readout architecture. Then its PLL bandwidth can then be made as small as required to suppress the phase noise without needing to move to a more expensive LO architecture.

Even neglecting the settling time of the local oscillator, one problem with rapidly sweeping or stepping the oscillator frequency is that the resulting excitation signal is not persistently exciting in the time domain. Therefore, an error will be introduced due to the transient response of the resonator. This is particularly true with high Q resonators. In order to mitigate this error, the time taken for each measurement point must be significantly longer than the resonator transient time, else it will affect the measurement accuracy, causing bias. The transient response of a resonator to a step change in driving frequency will decay according to $\exp(-\sigma_1 t)$, the same as for a step change in resonant frequency derived in the theory section. For example, a high Q resonator might have a bandwidth of 100 KHz. The settling time of this resonator to within an insignificant error of $-80\ dB$ is $t \approx -\ln(10^{-4})/\sigma_1 = 184\ \mu s$. This time must be multiplied by the number of sweep points to give the total measurement duration. In this example, a sweep of 100 points would be limited to a maximum repetition rate of about 50 Hz, an order of magnitude slower than that possible with the new prototype readout. However, when accuracy is not as critical, some of the bias introduced by the resonator transient can be cancelled out by first sweeping in the positive (increasing frequency) direction and then in the negative (decreasing frequency) direction; the resulting hysteresis is then ideally eliminated by averaging the two results.

The second option, attempting to keep the excitation frequency on resonance, is fraught with difficulty, as an additional frequency control loop is created which must compete with the existing phase locked loop (PLL) of the LO unless the tracking bandwidth is very small. A small tracking bandwidth limits the maximum

slew rate of resonant frequency possible for dynamic behaviour and therefore offers no advantage over the first option. Yet when the tracking bandwidth is increased to be comparable to the PLL bandwidth, it becomes impossible to separate a 'real' change in phase due to the resonance from that due to PLL lag. In addition, the accuracy in estimating resonant frequency decreases the further the excitation is from resonance.

Any feedback loop attempting to track dynamic changes in resonant frequency occurring at a rate close to or faster than its own tracking bandwidth may not converge reliably and may even diverge from resonance due to nonlinearities and bias in the estimation process. If this happens in an online measurement or control scenario, the results could be disastrous – in effect causing the MRS readout to 'loose lock' to the resonance. Manual intervention is likely necessary to reset this. I observed this effect in early experiments to measure dynamic effects such as heating with both a VNA and the early prototype MRS readout. When feeding back resonant frequency estimates, *e.g.,* derived from curve fitting, for example, errors in resonant frequency when a fast perturbation occurs cause divergence of the frequency control loop, especially as the error is higher the further the resonance is from the centre of the sweep. This was a key reason for keeping the LO frequency fixed in the prototype MRS readout and for removing resonant frequency tracking from swept VNA measurements in later experiments.

On the other hand, if tracking is to be avoided, then the problem with keeping the excitation frequency fixed in CW readout (option 3) is the reduced perturbation the system can reliably measure – in effect reducing its dynamic range. Furthermore, if a sample or environmental variation causes the resonance to drift outside of this narrow measurement range (no more than a fraction of its bandwidth) then without tracking it is necessary to do another swept characterisation in order to restore the excitation frequency accurately on resonance. This requires the suspension of measurements and the removal of the resonator from the dynamic environment it is measuring (option 4). In a realtime application, this may not be physically possible. Even if it is feasible, the system may miss some important temporal event during this calibration phase that it is

attempting to monitor, such as a change in composition or a flow system disturbance.

## 4.4 Prototype MRS readout system design

To address better these issues, the solution lay in the excitation of the resonator at multiple frequencies *simultaneously*. In this way, spectral characterisation over a sufficiently wide band is possible without sweeping the local oscillator. A persistently exciting signal with multi-frequency content also avoids resonator transient errors because it has no time-domain discontinuity. However, unlike CW excitation, permits accurate and unambiguous recovery of resonant frequency and bandwidth information without requiring periodic swept calibration. It can therefore make high rate MRS measurements over an unlimited period without concerns about system drift.

However, what hardware architecture permits the generation of a broadband excitation signal consisting of a number of discrete frequencies simultaneously? Clearly, having multiple RF synthesisers summed together with a power combiner is impractical on cost grounds. The answer, as alluded to earlier, lay in the technology of high bandwidth software defined radio systems. The community of researchers working on microwave kinetic inductance detectors soon realised that these systems were ideal for generating such multitone signals [29]. But whereas these multiple tones were used to address each individual resonator 'pixel' with a separate single tone, the same system architecture could be adapted to provide a multifrequency excitation to a single resonance instead, providing an ideal platform for a realtime, potentially low cost and portable MRS system.

I designed a new hardware architecture based around a direct-conversion, or 'zero-IF' wideband radio transmitter and receiver. Whilst direction-conversion architectures are known to suffer from problems not encountered in multi-stage analogue-based IF receivers (IQ mismatch, even order distortion, DC-offset and LO phase noise feedthrough [30]) they are simpler and require only one local oscillator. Furthermore, digital IF conversion, as used in modern VNAs, can still be implemented using the same hardware, allowing the system extra flexibility.

The system is based on quadrature mixers for upconversion and downconversion, but whereas in the homodyne quadrature system these are necessary to recover the amplitude and phase response of the resonator, in this radio system they are used to provide upper and lower sideband separation, or single sideband (SSB) modulation and demodulation. This allowed the bandwidth of the system to be utilised in the most efficient way. The maximum measurement bandwidth of the system is equal to the sampling frequency of the ADC and DACs used, or 100 MHz in this prototype. This is enough to measure even low Q resonant sensors under all but the largest sample perturbation conditions without needing to change the LO frequency. SSB modulation comes at a price; it effectively doubles the number of baseband components needed, requiring two ADCs and two DACs instead of one, as well as an IQ mixer at both the transmitter and receiver.

SSB modulation is based on a simple mathematical axiom that the Fourier transform of an analytic signal has no negative frequency components. Because negative baseband frequencies become lower sidebands and positive frequencies upper sidebands, forming the analytic representation of a signal before upconverting it should cancel the lower sideband entirely. Similarly, reversing the process for downconversion will reject any signals present in the lower sideband from being imaged into the positive baseband frequencies. The analytic representation of a signal is $z_a(t) = z(t) - j\mathcal{H}(z(t))$, where $\mathcal{H}(.)$ is the Hilbert transform.

IQ mixers approximate the upconversion of a complex signal, as noted in the previous section, so all that remains is to generate the real and imaginary parts of this baseband analytic signal. While this can be done with a hybrid, creating an analogue SSB modulator, this is necessarily a bandpass component and suffers from the imperfections of physical hybrid circuits. However, by generating the real and imaginary parts of the baseband signal with separate DAC channels, not only can the analytic representation be generated perfectly, independent upper and lower sidebands can be created, doubling the Nyquist band of frequencies that can be generated and received. This is why the bandwidth of the system is $f_s$ and not $f_s/2$ as it would be for a single ADC or DAC. A multitone excitation signal can thus

utilise both the negative and positive frequency components of the complex baseband signal, obtained easily in software by using the complex FFT. In fact, the vast majority of all the signal processing and manipulation can be done in the complex frequency domain, as discussed in the Chapter 5.

One elegant feature of this approach, despite the increased complexity, is that (with proper choice of excitation signal) the system can still recover the upper and lower sidebands perfectly even in the presence of IQ mixer phase mismatch. IQ mismatch causes part of the upper sideband to be reflected into the lower sideband and vice-versa, as shown in Figure 2.13.

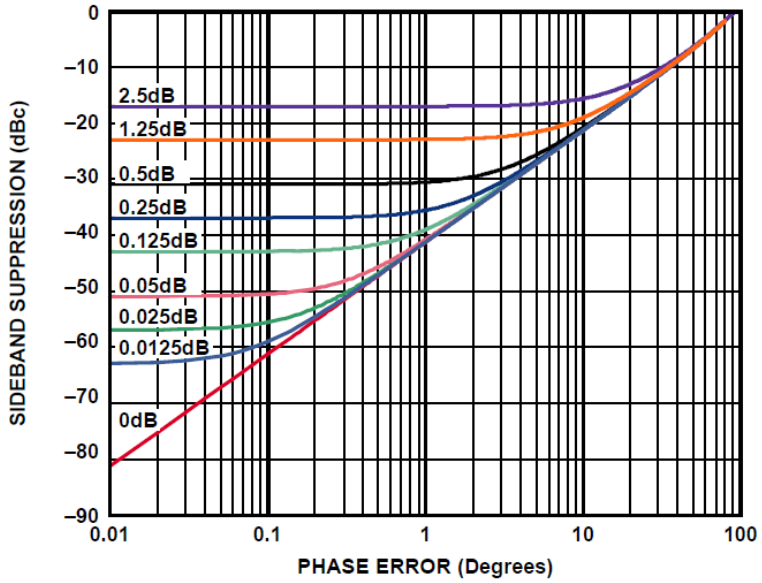



*Figure 4.13: Sideband suppression vs. quadrature phase error and various gain errors (shown by different colours with dB offset labels), from [31].*

As shown in Figure 4.14, by spacing the tones of the excitation asymmetrically about DC, or by digitally modulating the excitation around an offset frequency ($f_s/4$ is particularly good as it also minimises distortion), the spurious signals caused by this mismatch do not fall on the intended signal frequencies and thus can be recovered or simply filtered away. Note that tones are avoided close to DC – this bypasses the issue with local oscillator feedthrough and downconverted phase noise at this point.

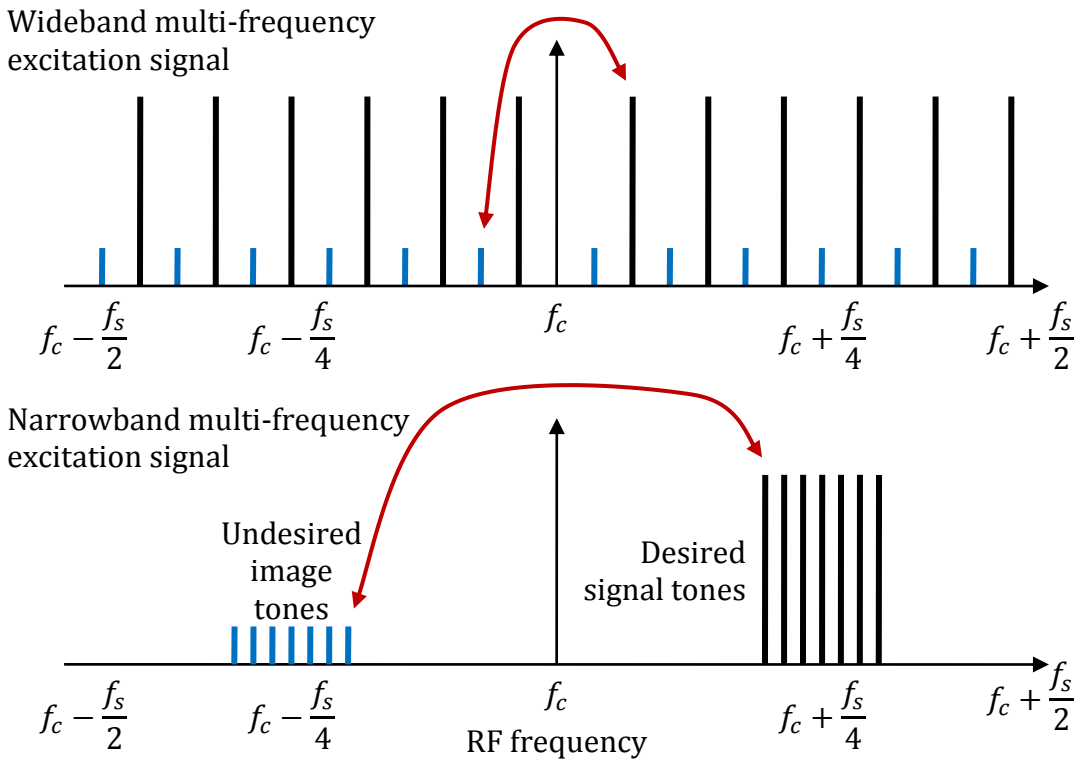



*Figure 4.14: Cartoon showing the relative magnitude of multisine tones transmitted subject to quadrature imbalance and imperfect sideband suppression. A wideband DC-centred signal with offset upper and lower sideband tones (top) and a narrow band digitally modulated signal (bottom) both allow the receiver to remove the effect of sideband images from the received signal.*

Figure 4.16 below shows a typical signal spectrum received by the implemented prototype system after passing through a transmission resonance. Since the resonator has quite a narrow bandwidth, digital modulation is used to centre the multitone signal around one quarter sampling frequency. Due to the quadrature receiver, the upper sideband image can be identified as a reflection of the main signal at negative baseband frequency. The fifteen measurement tones are clearly identifiable and can be selectively filtered in the frequency domain, simply by a 'brick wall' style filter. This is only possible because the tones were designed to be periodic in the 256-sample window. There is therefore zero spectral leakage and the excitation signal tone energy is confined perfectly to one FFT bin.

Careful attention to these issues, as well as the use of double balanced IQ mixers and proper selection of the LO drive and RF levels, mean that this prototype system can recover the resonator response with a very high degree of accuracy. In fact the only distortion to the power spectrum is due to odd-order intermodulation (IMD), which can be seen causing the spectral regrowth 'skirts' visible at around $-70$ dBFS to each side of the main multitone signal. This implies that the dynamic range of the spectral response measurement in this example varies from about 40 to over 50 dB across the resonator response. Whilst this is incomparable to a VNA's 90+ dB of dynamic range, it is certainly low enough to cause minimal systematic error in resonator measurements (a topic analysed in more detail in Section 5.1.1), and is either superior to or comparable with other, non-VNA MRS readout technologies.

Removing components such as couplers, switches, *etc.,* from the signal path, the complex 12-term error model of a VNA can be reduced to the simple error model discussed in Section 3.3.7, reproduced in Figure 4.15.

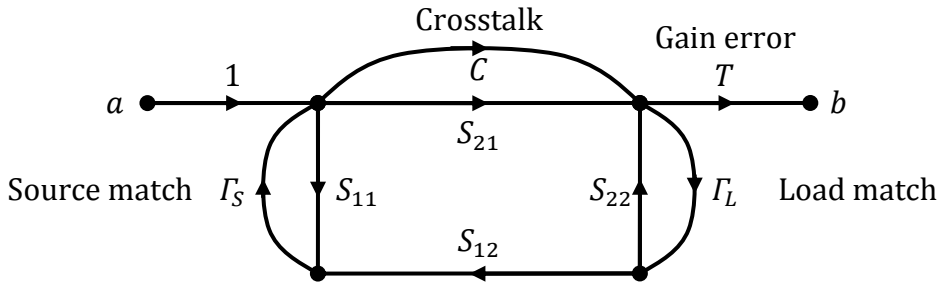



*Figure 4.15: Error flow graph of the multitone system.*

If the source and load impedances connected to the resonant sensor are very well matched (quite possible within a narrow frequency range such as the operational span of a resonant sensor) the only significant error term remaining is the complex gain term, $T$ (transmission tracking error). This can readily be calibrated for by connecting a 'thru' line and performing a basic response calibration, without affecting the accuracy the S21 estimate made by the system. Thus, the complexity

and labour required of VNA calibration are eliminated in this system without unduly introducing error.

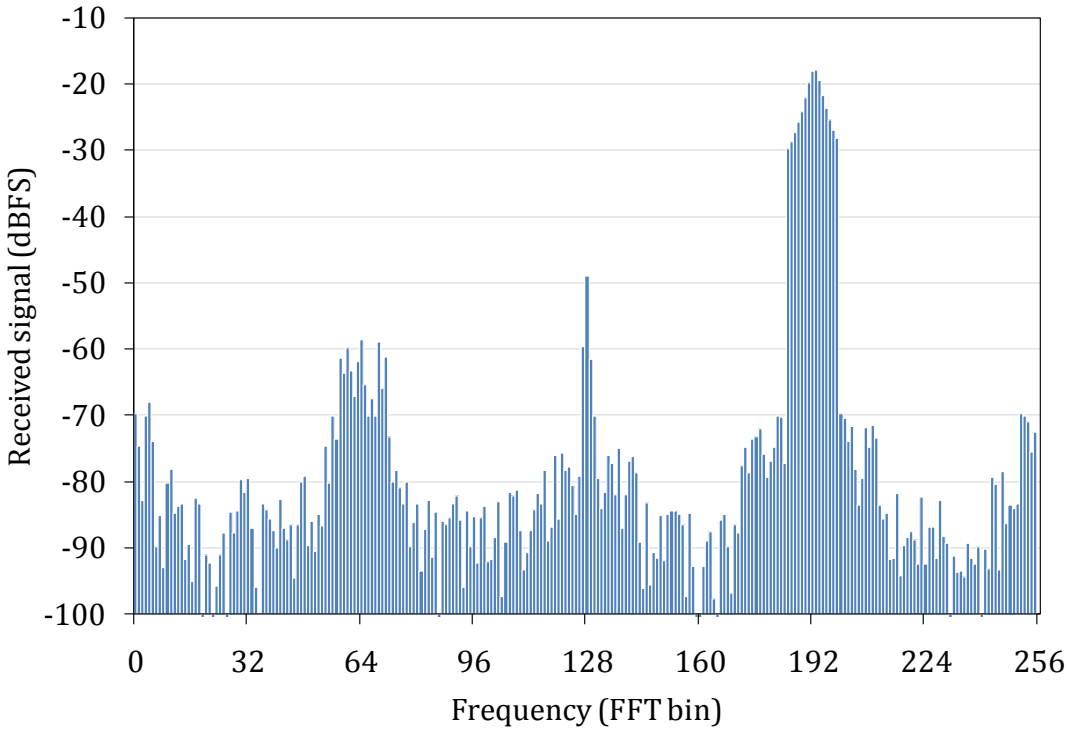



*Figure 4.16: Typical measured spectral response of the prototype system when measuring a resonator using a digitally modulated multitone excitation signal.*

The design of the hardware prototype is shown in Figure 4.17. The design has a number of features to ensure full coherence of the excitation and response signals. The transmitter and receiver share the same LO for coherency. Furthermore, all the converters (ADCs and DACs) share the same sample clock. This means they sample simultaneously, allowing the I and Q channels to be perfectly phase-synchronous. Digital triggering is also used to ensure the ADC records are aligned with the transmitted waveforms, allowing for accurate phase response estimation. In the system shown, amplification is done solely at baseband where low noise amplifiers are cheaper and have better performance, although in the experimental system RF LNAs were used at the transmitter because they were more readily available. Due to the poor VSWR of the double balanced IQ mixers, isolators are necessary to ensure a good source and load match.

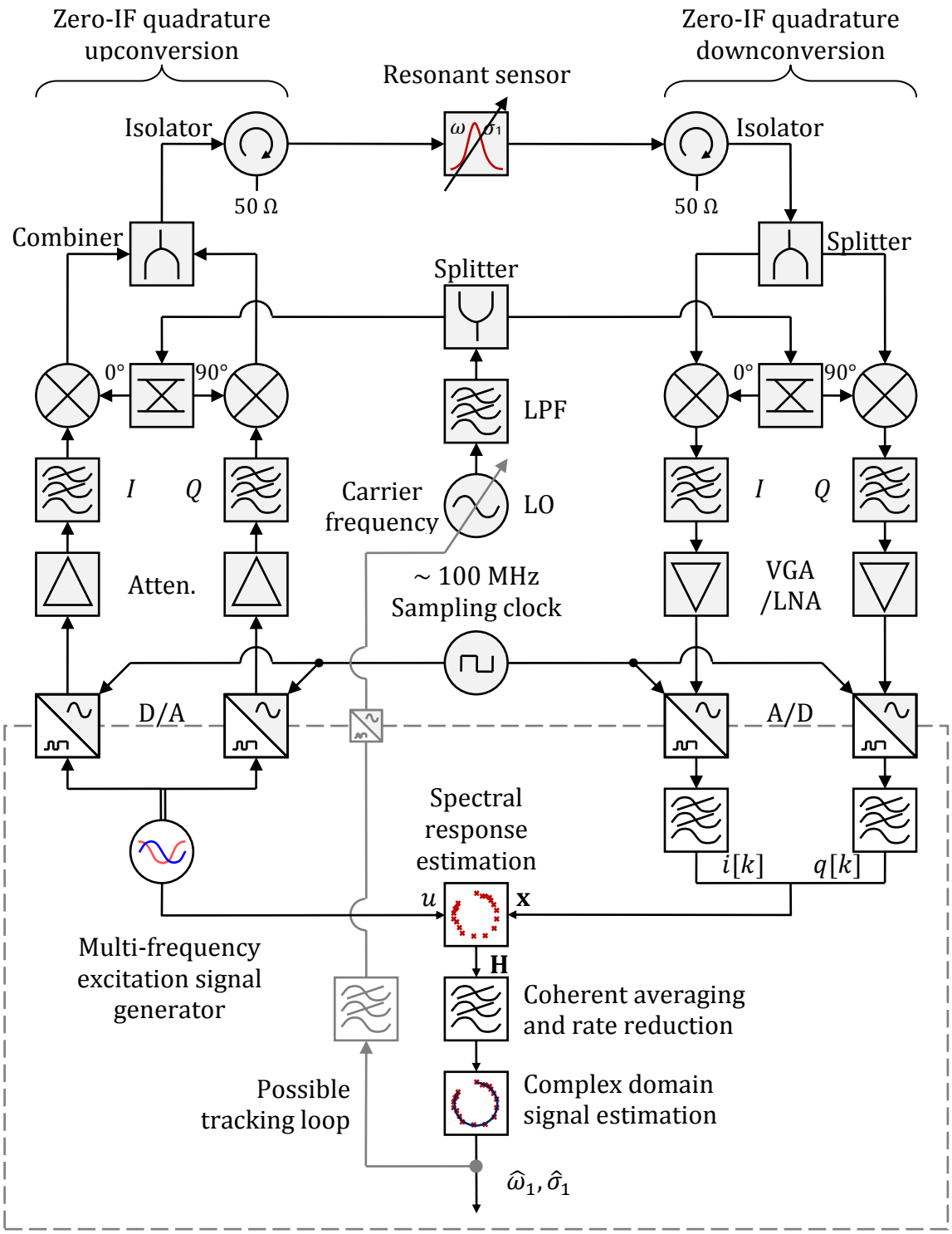



*Figure 4.17: Design for a simultaneous multitone excitation based MRS readout system based on a wideband coherent quadrature zero-IF architecture.*

Although in communication systems QAM requires IQ mixing to be used in order to recover the modulated phase and magnitude, this is because the LO at the transmitter and receiver are independent and therefore cannot be perfectly phase synchronous. However, in this system the same LO is used at the transmitter and receiver – thus quadrature mixing is not absolutely necessary and a single mixer could still recover the phase and magnitude response (just like in a VNA). In this system, a numerically controlled oscillator (NCO) is shown in the digital section as a means to perform IQ mixing numerically – however in practice FFT analysis would be used directly here as well, bypassing the need for this as an explicit step in the processing. Figure 4.19 shows a design for a simplified system that uses this approach and relies on the bandpass response of the resonant sensor to remove the lower sideband image from the signal present at the receiver.

As discussed in Pozar, [18], a balanced mixer that uses a 90° hybrid at its inputs instead of the usual 180° hybrid can ideally give a perfect input match. This is ideal for this simplified (potentially low cost and portable) system as it eliminates the requirement for isolators to ensure good source and load match conditions – an otherwise relatively large and expensive component.

As it stands, this simplified system has no bulky microwave transmission line circuits such as hybrids and couplers. It also has no active RF components apart from the LO (if passive mixers are used). It could therefore be realised in an extremely compact form without resorting to monolithic microwave integrated circuit (MMIC) technology. The downside to this, as indicated in Figure 4.18, is that incomplete lower sideband separation would cause the spectral error to be increased – degrading overall measurement accuracy.

Note that both these systems are inherently wideband. With a wide tuning range LO, they can be of almost unlimited frequency bandwidth. Whilst very wide bandwidth mixers are expensive, a flexible and accurate system such as this would not be specific to a particular MRS or frequency band. It therefore could be an attractive alternative to a conventional VNA for traditional laboratory MRS

applications, whilst at the same time enabling high-speed realtime readout capability.

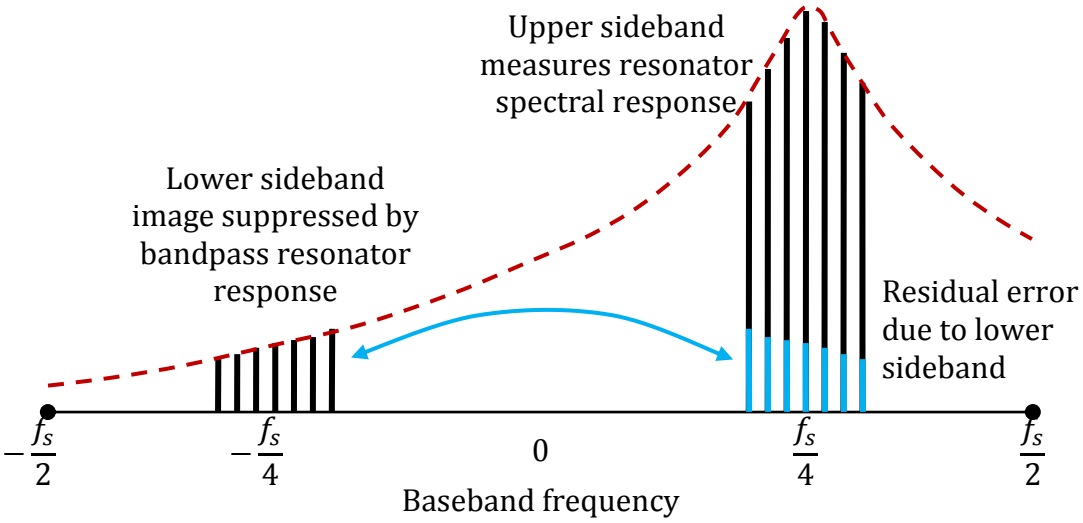



*Figure 4.18: Cartoon showing the relative magnitude of multisine tones received by the reduced complexity system. In the reduced complexity system, having a single channel ADC and DAC and no quadrature mixing, rejection of the lower sideband is performed by the bandpass response of the resonator. If the resonator attenuation of the lower sideband is insufficient, the remaining lower sideband signal is reflected back into the upper sideband at the receiver. Depending on the relative phases of the tones this interference may be additive or destructive, causing systematic error.*

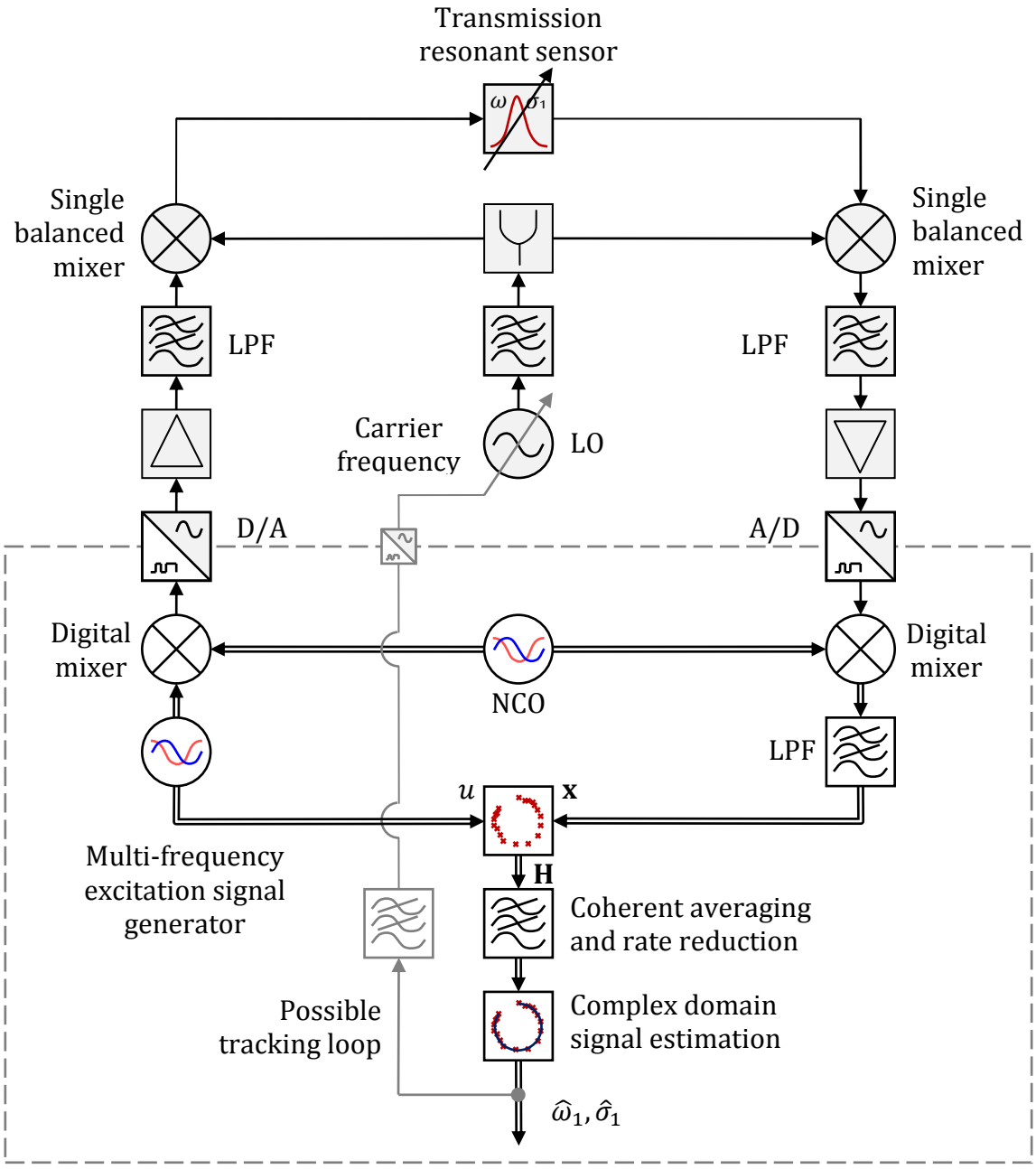



*Figure 4.19: Design for a simplified readout system that uses digital IF modulation and relies on the bandpass resonator response to provide lower sideband image rejection. Having no inherently bulky or expensive components, it is ideal for low cost and portable MRS instrumentation.*

Radio systems such as these can readily be implemented on a single chip at RF using modern silicon germanium (SiGe) processes, and the technology promises to be applicable throughout the microwave band and even into mm waves [32].

Several fully integrated quadrature receiver and transmitter chips have been developed by the high-volume manufacturers such as Analogue Devices (see Figure 4.20 below). The technology of this MRS readout system is ideally suited to full integration on-chip, making the ultimate goal of widespread, low cost, and low power consumption microwave sensor technology seem a step closer.

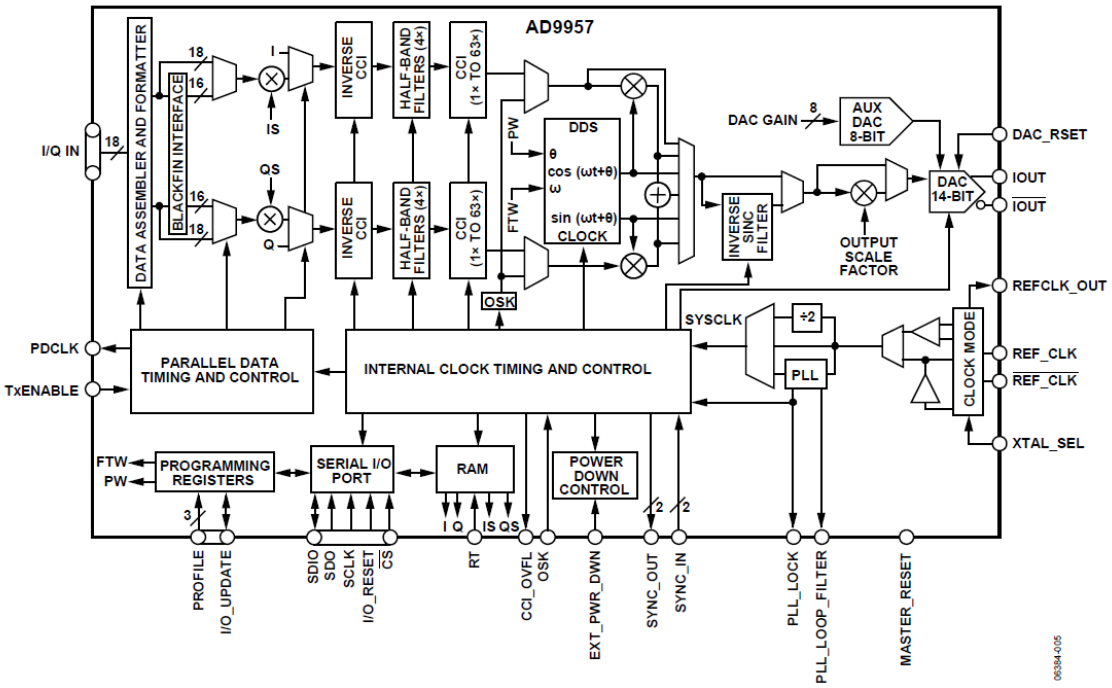



*Figure 4.20: Functional block diagram of a state-of-the-art, fully integrated single chip quadrature digital upconverter with 14 bit, 1 GS/s DAC [33]. The 100-pin 16 mm square integrated circuit has a typical power consumption of 1.4 W.*

## 4.5 System implementation

These system designs are still at a high-level of abstraction; the design of a physical system to implement them from scratch at the component level would be a significant undertaking and require a degree of time and capital investment no less than the development of any advanced instrument – something that was not available at the time of doing this work. A way of realising them using existing hardware and instruments as far as possible was therefore required. This meant that as much time as possible could be devoted to the novel aspects of this work such as the system architecture, algorithms and calibration methods, rather than on implementing already existing technology.

To this end, I initially used an existing instrument set that emulated the entire transmitter and receiver hardware. Having excellent performance and specifications, these instruments provided an ideal starting point to develop algorithms and perform some initial experiments. However, this system was both expensive and not fully suitable; it lacked perfect coherency between the transmitter and receiver because they were unable to share the same LO.

The subsequent implementation used instruments to provide only the baseband ADC and DAC functionality, as well as a synthesiser to implement the LO. This allowed the RF system to be developed at the component level whilst using the previous hardware as a useful performance reference. These instruments were provided for by a £20,000 grant from the Cardiff Partnership Fund (CPF), which was awarded based on the commercial potential of this technology. As part of this project, a patent application was written detailing the system and its key algorithms for realtime operation. At the time of writing, this patent has been filed and is awaiting examination [34]. The subsequent subsections discuss these two hardware implementations in more detail.

### 4.5.1 First implementation using VSA/VSG

I developed many of the working principles, software algorithms and performed some of the earlier experiments, such as dynamic microwave heating, using a National Instruments (NI) vector signal generator (VSG) and vector signal analyser (VSA). The VSA was the NI PXI-5661 model and the VSG used was the NI PXI-5670. The devices covered an RF range of 250 KHz – 2.7 GHz.

The VSA uses a superheterodyne IF architecture and digital IQ downconversion using a single channel of a 14 bit, 100 MS/s digitizer. The superheterodyne architecture allows it to eliminate sideband images and carrier feed through. An expensive and delicate yttrium iron garnet (YIG) preselection filter, tuned magnetically, provides high out-of-band and noise rejection at the input. It also has very low input distortion and low a noise figure due to high quality passive mixers and LNAs. Like most RF instruments it relies on a very low phase noise synthesiser referenced to a 10 MHz oven controlled crystal oscillator (OXCO). An OXCO holds a quartz crystal at a turning point in its temperature-frequency curve, well above external ambient temperature, using a well-insulated oven with thermostatic control loop. This provides an accurate PLL reference that ensures low LO phase noise below 10 MHz.

The VSG used a very similar architecture in reverse, a 16-bit 100 MS/s arbitrary waveform generator (AWG) with digital upconversion is followed by superheterodyne IF upconversion. 'AWG' is term used to describe a system containing a DAC plus some supporting functionality, generally variable output amplification, reconstruction filters, *etc.,* along with a digital triggering and memory core that generally supports a range of built-in waveform patterns. Again this multi-stage upconversion system is extremely high performance and uses a tuned YIG filter for harmonic, spurious and noise rejection. The VSA also contains its own high performance LO.

These instruments together therefore provided an excellent reference – or emulator – of the required receiver and transmitter functionality, approximating

the performance of an 'ideal' radio system from a comparative perspective. Although in some ways, their architecture is similar to the direct conversion design, the exclusive use of digital upconversion and superheterodyne IF architecture means that sideband suppression, LO feed through suppression and IQ match are essentially perfect. However, because digital upconversion makes incomplete use of the full ADC Nyquist range, both the VSA and VSG have only 20 MHz realtime bandwidth, which is further limited to only 8.3̇ MHz when using the onboard signal processing for digital upconversion in the AWG. Furthermore, at a total cost of around £28,000, this system is clearly not the basis of any sort of competitive technology to the conventional VNA for MRS readouts.


*Figure 4.21: The NI PXI-5670 VSG (left) and the NI PXI-5661 VSA (right) shown withdrawn from the PXI chassis where they are normally mounted [35].*

The instruments are mounted in a chassis that provides their interface, power and ventilation requirements. Because they have no display or controls of their own, they are known as virtual instruments (VIs), as a user interface must be implemented in software on a connected PC. The communication bus used in this chassis is known as PXI – hence the devices and chassis are known generally as 'PXI instruments'. The peripheral component interconnect (PCI), extensions for instrumentation (PXI) form factor, first introduced and standardised by NI in 1998, has several major advantages for implementing realtime systems [36]. Firstly,

being based on the PCI standard which supports a peak data rate of 133 MB/s, it allows very rapid transfer of large amounts of captured data from the digitizer to be streamed to a PC for analysis in realtime. Without instruments based on this form factor, or its even faster successor PXI express, (PXIe) that supports a peak transfer rate of > 1GB/s, implementing this experimental platform for realtime resonator measurements would have been impossible. By comparison, the former *de facto* standard general-purpose instrumentation bus (GPIB), IEEE-488 has a peak data transfer rate of just 1.8 MB/s. The use of this instrument form factor and acceptance of the virtual instrument concept was therefore essential to the overall success of the hardware prototype.

In order to synchronise the instruments properly the signal routing shown in Figure 4.22 was used. Synchronisation was done both at baseband and at RF. As the VSG and VSA each had separate LOs, and neither instrument supported an externally supplied local oscillator, the best RF synchronisation that could be achieved was by exporting the 10 MHz reference from the VSG to the downconverter. Whilst this ensures the frequencies of the local oscillators are locked, it does not guarantee phase coherency. Therefore, a digital signal processing approach was necessary to individually phase synchronise the received waveforms with the excitation at baseband. To synchronise the baseband, the sampling clock of the AWG is exported to the digitizer (ADC), which is programmed to accept this externally supplied clock for data conversion. In addition, the AWG is set to output a pulse every cycle of the periodic excitation signal (using the AWG 'script' mode and marker output functionality). This used to trigger to each digitizer acquisition. Although there is a fair amount of trigger jitter, using this configuration the baseband generated and received waveforms are still guaranteed to be synchronous to within less than one sample period, on average.

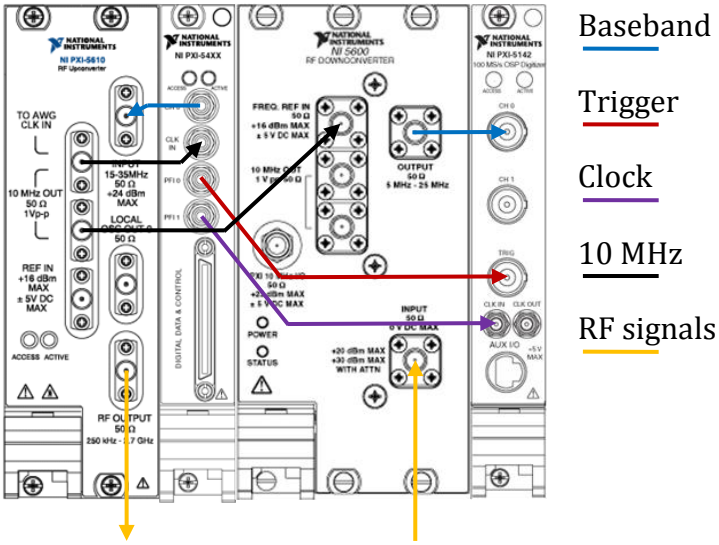



*Figure 4.22: Front panel signal routing used to synchronise VSG and VSA.*

### *Second implementation using custom RF front end*

The VSA/VSG system was functional, but it suffered from performance issues due to the lack of coherency – meaning the system could not take full advantage of these instruments' low distortion and noise performance. Furthermore, the system was still a long way from showing that the multitone hardware architecture could be realised in an economical and efficient way. The CPF funding was used to purchase two NI AWGs (PXI-5422), an NI digitizer (PXI-5122) and an RF signal generator (RFSG) (PXI-5652) in order to implement a second hardware prototype.

The AWGs are single channel, 16 bit, 200 MS/s DACs that include analogue filtering and gain control, as well as a synchronisation and memory core (SMC) providing waveform memory and scripting functions, trigger and clock routing, as well as the same high speed PXI interface. They thus provided I and Q channel DAC functionality without necessitating difficult and time consuming digital design or FPGA programming. Their LabVIEW drivers provided all the necessary functionality to configure and send arbitrary data and configure relevant triggers, *etc.* The digitizer is a 14 bit 100 MS/s two-channel ADC with input amplification, filtering and buffering and well as the same SMC and similar functionality. The RF synthesiser has a range of 500 KHz – 6.6 GHz and thus provides plenty of flexibility

for testing the system with a range of different resonant sensors. This was a major improvement over the limited 2.7 GHz upper frequency of the previous implementation.

Initially, an RF front end based on active IQ modulators and demodulators; the Analogue Devices (AD) ADL5375 400 MHz-6 GHz evaluation board and the ADL5380 quadrature demodulator evaluation board. Additionally, in order to drive the balanced inputs of the modulator, two broadband, low-noise, low-distortion differential amplifiers (Texas Instruments THS4509 also as evaluation boards) were used rather than baluns to ensure full modulation bandwidth. Whilst these components are miniaturised and have good performance characteristics on paper, they were not ideally suited to implement the RF front end. Firstly, due to sensitivity to DC input bias variations and temperature, the performance of these devices was prone to drift. This made it difficult to apply simple predistortion methods from a one-off calibration. In addition, the modulator had poor noise performance at high frequencies (noise figure 15.5 dB at 5.8 GHz). Tricky manual circuit modifications were also required to alter RF balun for different MRS frequency ranges. These issues, and the generally time consuming setup (requiring numerous power supplies and DC offset nulling) meant that the IC solution was dropped in favour of simpler passive IQ mixers.

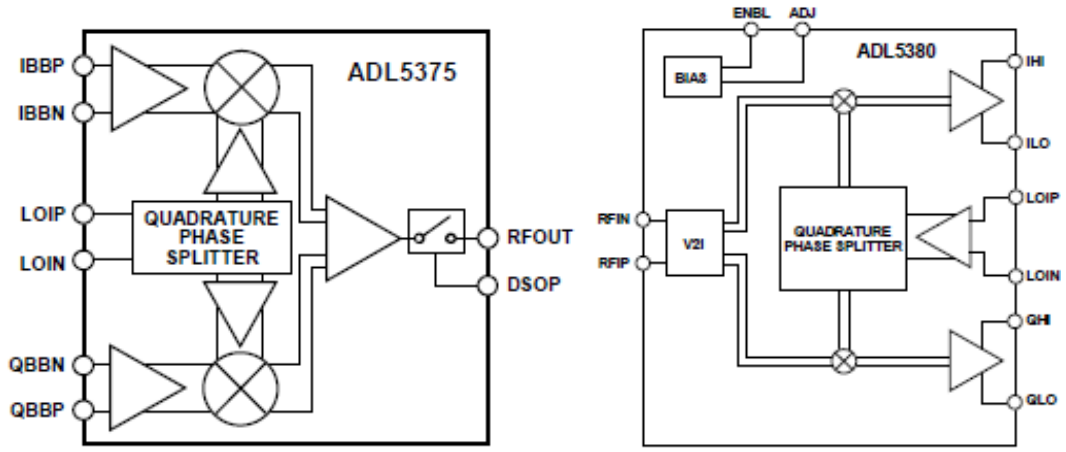



*Figure 4.23: Functional representation of the ADL5375 active IQ modulator and ADL5380 active IQ demodulator [37], [38].*

Identical Marki microwave double balanced passive quadrature mixers were used in the final implementation of the system (IQ-1545LMP, Marki Microwave Inc., Calif., USA). These mixers had a frequency range of 1.5 – 4.5 GHz, suitable for the range of MRSs available for experimental work. The models ordered specified a LO drive level of 10 dBm, but were found to perform almost identically with as little as 6 dBm whilst lowering the output requirements of the RFSG and reducing the LO feedthrough at the receiver. The performance specifications of these mixers was very good (see below), but at £380 each they were not as cheap as solution as the AD active components. However, the simplicity of the RF front end setup and reliable, low-drift performance was a boon for experimental testing.

*Table 4.2: Performance specifications of the chosen Marki Microwave double balanced quadrature mixers [39].*

| Parameter | Minimum | Typical | Maximum |
|---|---|---|---|
| Conversion loss (dB) | | 5.5 | 7 |
| Image rejection (dB) | 16 | 25 | |
| I/Q gain mismatch (dB) | | 0.3 | |
| I/Q phase mismatch (degrees) | | 3 | |
| LO-RF isolation (dB) | 37 | 43 | |
| LO-IF isolation (dB) | | 30 | |
| RF-IF isolation | | 30 | |

| Parameter | Minimum | Typical | Maximum |
|---|---|---|---|
| Input 1 dB compression (dBm) | | 4 | |
| Input third order intercept (TOI) (dBm) | | 14 | |

In addition to the passive mixers, a resistive power splitter (Mini Circuits ZFRSC-123+, £60), which uses a slightly unusual design [40] to achieve about 20 dB of isolation at an insertion loss of 10 dB (a conventional splitter would have just 6 dB). Whilst this is not very power efficient, it is compact, provides sufficient drive level to the mixers over a very wide frequency range and the additional isolation was found to improve the signal distortion levels – probably by eliminating a parasitic RF signal path between the modulator and the demodulator LO ports. Additional filtering was also included, extra 50 MHz antialiasing LPFs (Mini Circuits SLP 50+ £28) as well as DC blocking capacitors (Mini Circuits BLK-89+ £12).

Matching bandpass filters were also used at the RFSG output and in the transmitter path to provide harmonic, spurious signal and noise rejection. Various Mini-Circuits bandpass filters were used for different resonant sensor frequency ranges (*e.g.,*VBFZ-4000-S+ £32). A 3dB pad was placed immediately after the RF output of the IQ modulator – this was found to improve signal distortion significantly (possibly by attenuating the baseband signal reflected by the filter or resonator). This was followed by one or two cascaded 18 dB gain LNAs (Mini-Circuits ZX60-3018+) depending on the insertion loss of the sensor. Finally, isolators were used on the outputs to improve source and load match (Aerotek Co., Thailand), with a specified VSWR of 1.4 these gave a noticeable improvement over the mixer VSWR of between 2-3, however, they could be eliminated by using a custom mixer designed for better RF match, as mentioned above. A photograph showing part of the RF front end and the PXI instruments is shown in Figure 4.24.

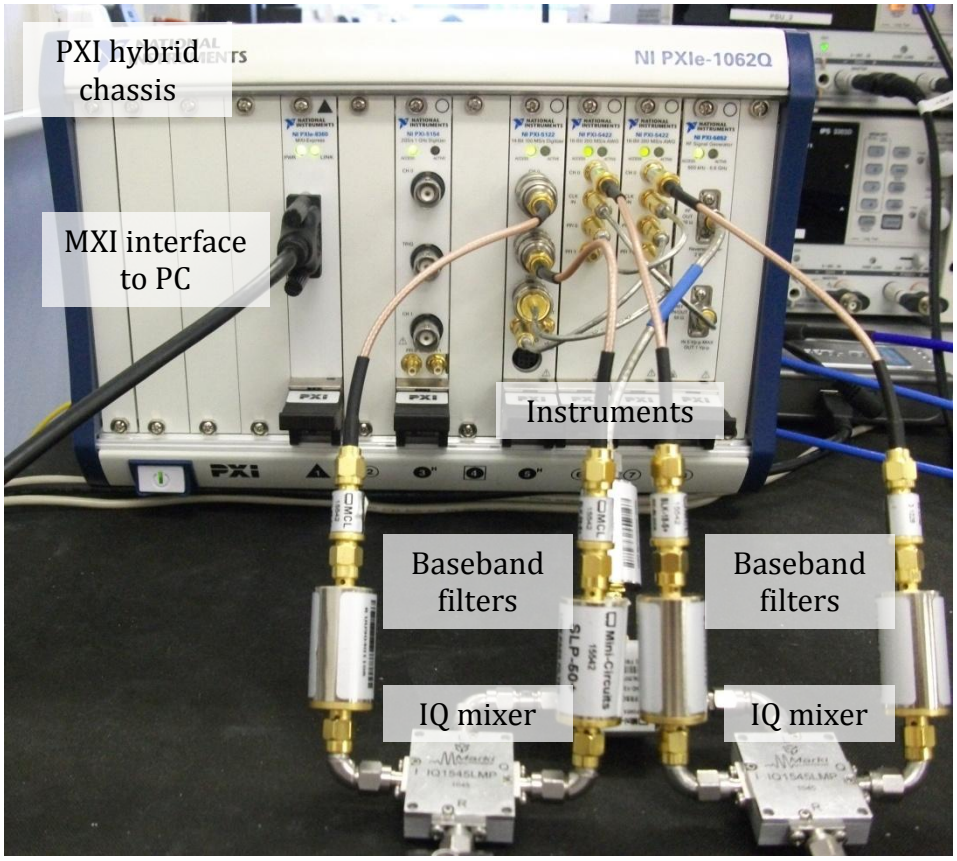



*Figure 4.24: Photograph of the second implementation of the multitone system using passive IQ mixers.*

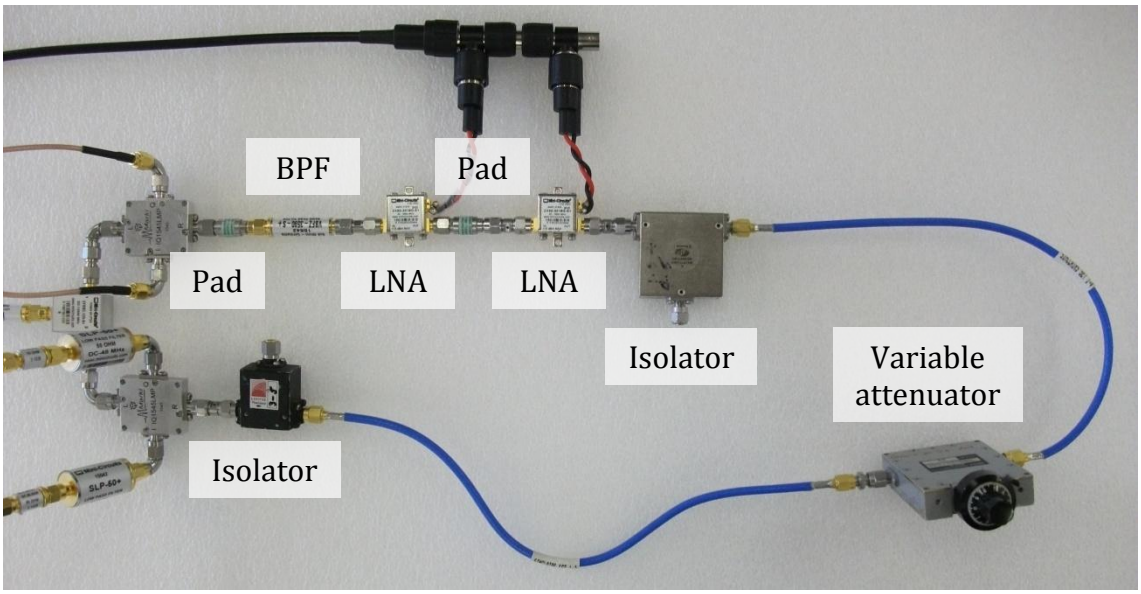



*Figure 4.25: Top-down view of the implemented RF front end with variable attenuator used for flatness and predistortion calibration. Two amplifiers are shown here cascaded with a 3 dB pad in-between to improve matching. As such this front-end is suitable for high insertion loss resonant sensors.*

Again, careful attention was paid to the ensuring coherency of the transmitter and receiver. Because the mixers were supplied with an identical LO signal, RF coherence was guaranteed. Both DACs and the two-channel ADC were supplied with the same 100 MHz sampling clock. The functionality to divide this clock signal frequency in a phase coherent way was not provided on the instruments, thus the AWGs were run at half their maximum sampling rate of 200 MS/s. The accurate and low-phase noise OCXO-derived 10 MHz reference signal from the RFSG was used to phase lock the sampling clock of the AWGs and ensure the sampling jitter was minimised. Finally, a waveform marker signal was used to trigger the waveform acquisitions in the digitizer. These measured ensured the transmitted and received signals were phase synchronous to with a fraction of a degree, much better than that possible with the previous system. Therefore, no additional software processing for synchronisation was necessary.

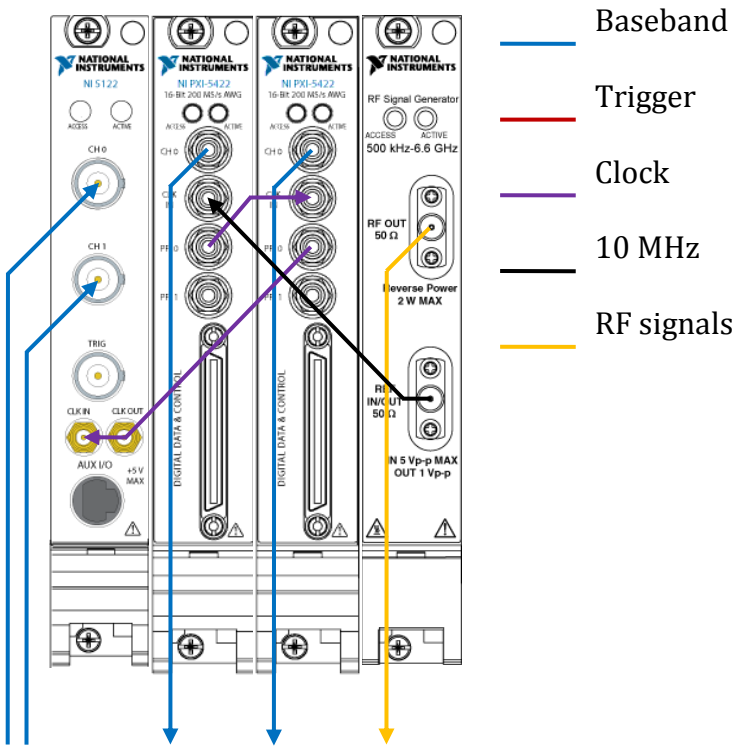



*Figure 4.26: Front panel signal routing to ensure full phase synchronisation between the baseband and RF instruments.*

The total noise figure of this system can be calculated by considering the cascade of all the RF system components. The combined IQ SNR at the output of the DAC can be estimated to be about 3 dB higher than that due to a single DAC's quantisation noise, assuming the noise adds in an uncorrelated way. The signal to quantisation noise (SQNR) is given by the following equation: $\mathrm{SQNR}_{\mathrm{dB}} \approx 6.02\,N + 4.77 - F_{\mathrm{dB}}$ where $N$ is the number of bits and $F_{dB}$ is the crest factor of the generated signal (assumed to have amplitude equal to the full-scale range (FSR) of the DAC). The noise figure of the system is given by the classic cascade equation:

$$F_{cas} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots . \qquad 4.8$$

For the implemented RF front end with single amplifier, the worst-case calculated noise figure is about 15.7 dB for a resonator insertion loss of 10 dB, and 14.4 dB for an insertion loss of 6 dB, when using a single 18 dB gain RF amplifier at the transmitter after the mixer. This is dependent on the conversion loss of the mixers,

although no single component is truly dominant in the calculation. This results in a combined SNR at the receiver ADC of about 79-81 dB for a signal with a crest factor close to a sine wave, and is thus comparable with the 14-bit ADC's own quantisation noise of 81 dB. However, the received signal will rarely be exactly equal to the FSR of the ADCs in the implemented system as there is no automatic gain control. In the system with one RF amplifier, the total path loss is 8 dB for a 10 dB insertion loss resonator. The maximum baseband amplification setting within the digitizer is 20 dB, so this path loss can be easily made up. Thus, the actual received signal SNR is typically dominated by the SQNR of the input ADCs. From this point of view, the implemented system can be considered to have more than adequate noise performance.

## 4.6 Optimisation of VNA measurements for realtime MRS readout

Although the above discussion criticised the VNA, in situations where cost, space, and power consumption are of no importance, and fast sampling (> 100 Hz) is not required, the VNA represents an extremely accurate and highly flexible, if somewhat redundant, platform for a MRS readout system. However, optimising a VNA for realtime MRS measurement is not a trivial task; these instruments are not designed with realtime measurements as a primary concern. In order to provide a fair performance comparison for the new hardware systems developed in this work, a parallel aim was to achieve the best possible performance out of existing mid-range VNAs. I developed procedures and control software to operate an Agilent E5071B and a Rohde and Schwarz ZVL, each costing approximately £20,000, in a mode suitable for dynamic MRS applications. Not only was this project valuable as a comparison, it has also proven useful for a number of other research projects that require high accuracy dynamic MRS measurements.

Optimising a VNA-based system means maximising the noise spectral density of the resonator measurements, *i.e.,* the total error per second or 'precision-per-point'. The MRS sampling period is the time between successive measurements of resonant frequency, permittivity, *etc*. By definition, 'realtime' means that processing must not be 'offline' but run continuously, as a minimum resonant frequency and bandwidth should be output at the intended measurement rate. Whilst most modern VNAs include very basic processing for resonant frequency and 3 dB bandwidth, this method based on discrete 'marker' interpolation performs poorly compared to curve fitting, as discussed in [41] (elaborated on in Section 5.3.1 which discusses curve fitting), especially as the number of sweep points decreases. Therefore, realtime VNA measurements require some additional processing on top of that performed by the VNA as standard. Furthermore, if this processing and data collection is performed on an external PC (the method chosen for all this work), there will also be a finite time required for the transfer of the VNA spectral data. Therefore, the total measurement period will be given by the

sum of the sweep time, internal processing time, data transfer time and external processing time.

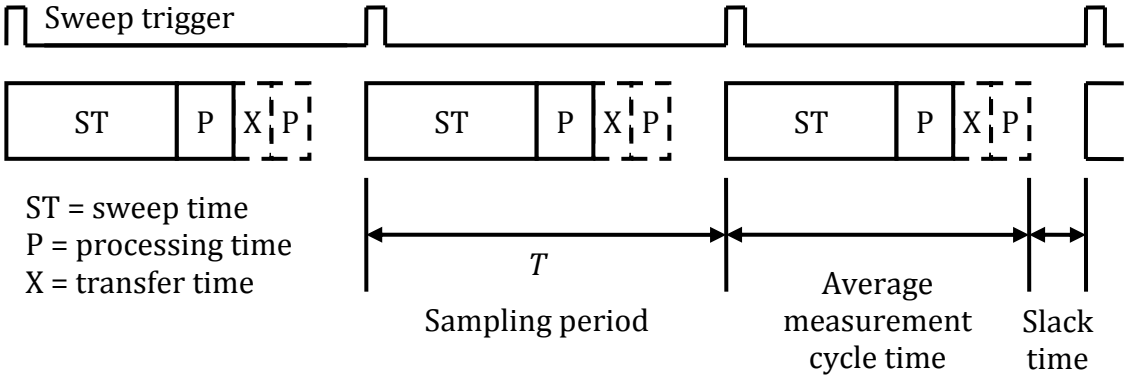



*Figure 4.27: Non-pipelined VNA measurement timing.*

Note that some slack time is necessary because all PC-based processing, as well as the data transfer, are nondeterministic and thus subject to some variability. This is one fundamental issue with VNA measurements, what can be thought of as sampling jitter. If VNA measurements are simply configured to run as fast as possible (by the VNAs internal timing) then the actual sample time when the resonance is measured is subject to a lot of jitter, something that was found experimentally. This is disastrous if important information is represented by the temporal axis. The capillary filling experiments, discussed in the Experimental Verification chapter, are an example of this. A partial solution that makes a big improvement to sampling jitter is to use an external sweep trigger, an external signal generator configured to output a square wave or regular pulse at a known frequency. In these and all subsequent experiments the VNA was triggered using a synthesised function generator (Agilent 33220A) set to output a repetitive trigger signal at a frequency controlled by the PC.

Another improvement in timing efficiency was made possible through adopting a pipelined architecture. Using LabVIEW's built-in multithreading capability allowed the process thread handling the data transfer from the VNA that to run in parallel with the data processing thread. This not only eliminated the additional processing time (*i.e.,* curve fitting) from the total measurement time, but also by further

concatenating several measurements together and processing them as a block, the impact of one-off overhead delays, such as the saving of data to file, could be minimised. The two loops were synchronised using a FIFO buffer (known as a queue structure in LabVIEW) allowing data to be transferred between threads without risking a software race hazard. Data streaming from the VNA was also handled in an innovative way. It was found that using the standard method to ascertain whether new data were ready (polling the VNA's status register) caused an undesirable overhead. Thus, the software used a 'quick and dirty' method instead. Data were streamed continuously at maximum rate and compared with the pervious iteration. Due to inevitable noise and time-variation, new data are invariably different from old, allowing the control program to establish that a sweep was complete and so add the new data to the buffer. These programming developments all helped improve the maximum readout rate and achieve truly the highest possible measurement rate.

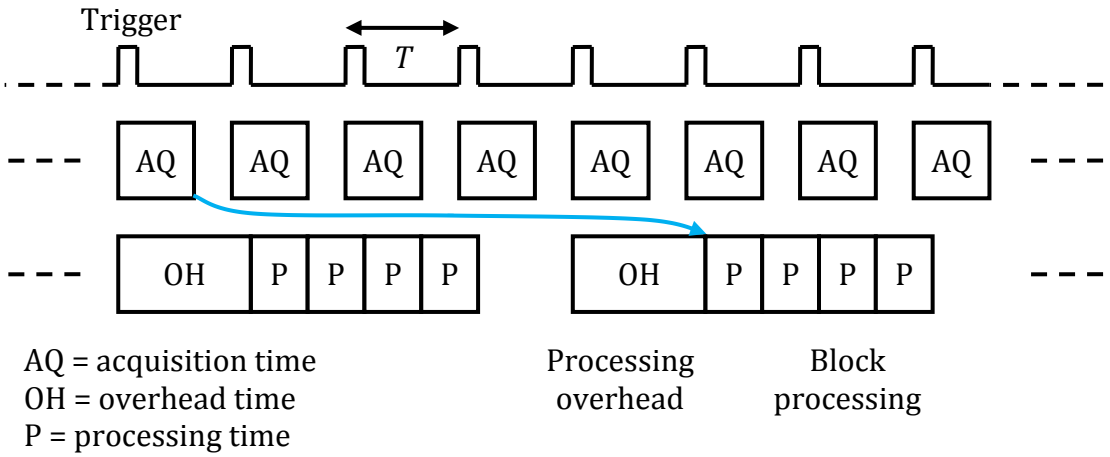



*Figure 4.28: Pipelined VNA measurement timing.*

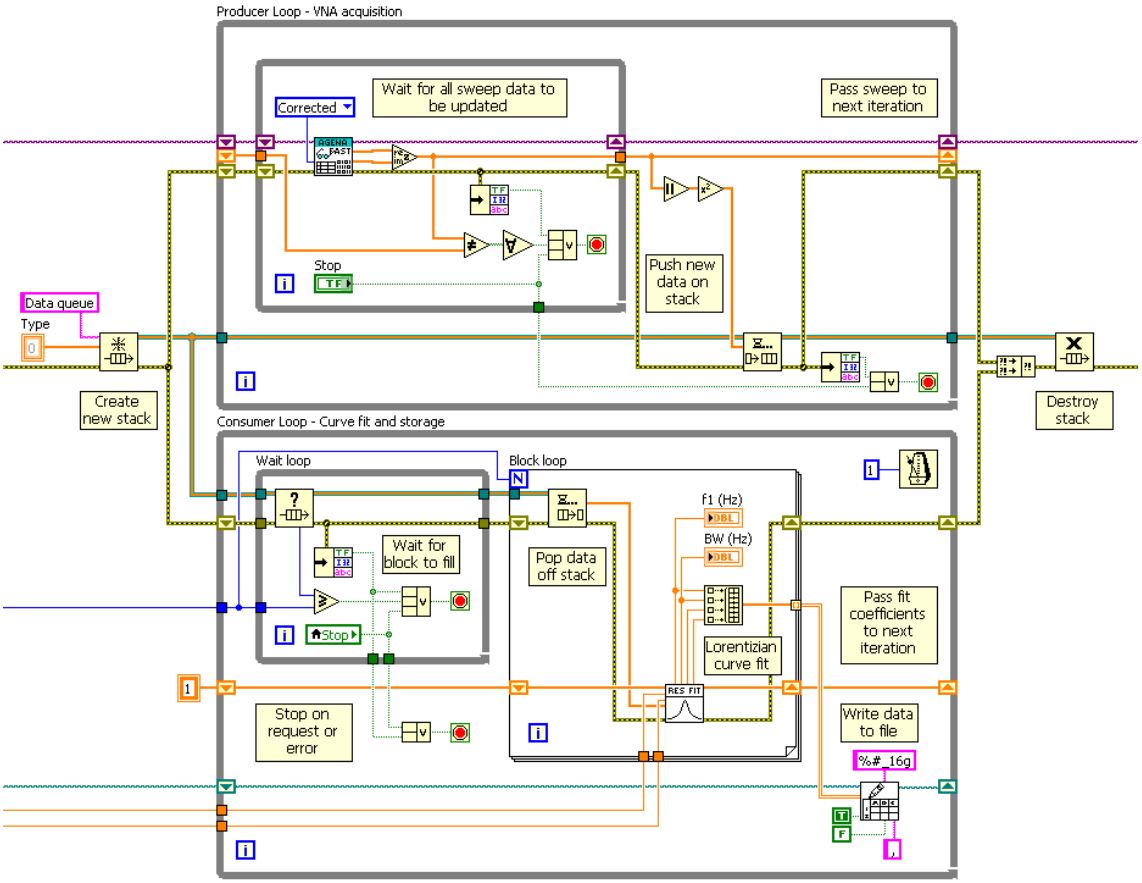



*Figure 4.29: LabVIEW virtual instrument (VI) for pipelined VNA realtime readout with curve fit processing and streaming to file.*

Due to these improvements, and the fast data transfer rate of the USB 2.0 control interface (the highest bandwidth instrumentation bus supported by the VNA), at all but the quickest sampling rates ($T$ > 50 Hz), the VNA sweep time dominates the total measurement period. To optimise temporal MRS measurements further, therefore, it is necessary to focus on the factors that determine the sweep time and how these affect the dynamic accuracy of measurements. In other words, for a given accuracy, what is the shortest sweep time (ST) that can be achieved?

The architecture of VNAs has been discussed in some detail above. The key parameter that determines the dynamic range of spectral measurements is the IF filter bandwidth (IFBW), also called the resolution bandwidth (drawing a comparison to spectrum analysers). On modern VNAs this is implemented by a digital filter which can be varied in bandwidth as desired with a simple software

change. Assuming the input has a white noise distribution (valid outside of limiting phase noise), *e.g.,* thermal noise spectral density, $\sigma^2 = 4Rk_BT$ $(W/Hz^{-1})$, and is terminated in a matched load, then the noise power density measured at the receiver, $P_n = k_BT$ $(W/Hz^{-1}) \approx -174$ dBmHz$^{-1}$ at 20°C. Thus, the SNR of a scattering parameter should be proportional to the IFBW. However, using a narrower IFBW increases the measurement time per point. Fundamentally, this is because of the relationship between time and frequency – the Heisenberg-Gabor limit [42]. This says that, for a given accuracy, a narrower bandwidth requires a longer 'integration time' and vice-versa; mathematically, $WT = U$, where $W$ is the bandwidth, $T$ is the sampling interval and $U$ is some uncertainty constant.

Thus to achieve a real reduction in noise by a factor of 10 not only implies a decrease in IFBW by a factor of 10 but also an increase in ST by the same factor. In practice, however, the digital filtering method is not perfect and has certain overheads associated with it. This means that the ST does not increase in proportion to decreasing IFBW but rather 'saturates'. Another way of looking at this is that a reduction in IFBW by 10 does not always increase ST by 10. This how manufacturers see it because it sounds better. Agilent provide a handy application note [43] that summarises these issues and the impact of IFBW and vector averaging on both noise floor and sweep time (see *Table 4.3*). This shows that, *from the noise floor point of view*, it is more time-efficient to reduce IFBW in preference to increasing vector averaging, however this advantage becomes insignificant towards the narrowest IFBWs. All modern VNAs should show a similar trend, although the exact figures will vary.

*Table 4.3: Comparison of noise reduction and sweep time for the Agilent PNA, ordered by increasing sweep time [43].*

| IFBW | Averages | Noise floor reduction | Relative sweep time |
|---|---|---|---|
| 10 KHz | 0 | 0 dB | 1 |
| 1 KHz | 0 | 10 dB | 7.75 |
| 10 KHz | 10 | 10 dB | 10 |
| 100 Hz | 0 | 20 dB | 74.8 |

| IFBW | Averages | Noise floor reduction | Relative sweep time |
|---|---|---|---|
| 10 KHz | 100 | 20 dB | 100 |
| 10 Hz | 0 | 30 dB | 740 |
| 100 Hz | 10 | 30 dB | 748 |
| 1 Hz | 0 | 40 dB | 7443 |
| 100 Hz | 100 | 40 dB | 7480 |

This answers one part of the question – avoid averaging for preference over IFBW. But how to choose the best number of points for a given accuracy? To get a better practical understanding of this, this fast VNA setup was used to measure a nominally static cylindrical $TM_{010}$ mode cavity resonator held at approximately 35°C using a warmplate (Minitube HT50). Different numbers of sweep point and IFBW were used and Lorentzian curve fitting was used to estimate resonant frequency and bandwidth in the realtime LabVIEW control program. The VNA was triggered by 1 KHz square wave, effectively instantaneously. The inter-measurement delay was measured using a software millisecond timer and, when averaged, gives a fair approximation to the maximum possible measurement rate under these settings. Recordings of resonant frequency and bandwidth deviation were made over a 1000 sample window. In order to assess measurement repeatability (precision), the average absolute forward finite difference (*i.e.,* the point-to-point deviation (PPD)) was calculated over the 1000-sample set for each IFBW-points combination. This statistic was chosen rather than the variance, absolute deviation, *etc.,* as the derivative minimises sensitivity to natural drift due to temperature, *etc.,* whist still giving a good practical measure of variability:

$$\overline{PPD} = \frac{1}{n-1}\sum_{k=2}^{n}|x(k)-x(k-1)| \qquad 4.9$$

The measurement speed results show that there is a significant overhead (probably the data transfer to the PC over USB) that causes the period to bottom out at around 6.5 ms. The fastest possible measurement rate with this system is therefore around 150 Hz. The period decreases slowly, crossing the 10ms threshold diagonally across the IFBW-points plane, as shown in Figure 4.30. At higher IFBWs, the measurement period increases roughly with its square, as expected. However, the measurement period does increase linearly with the number of sweep points, again probably due to overheads in instigating the sweep and establishing communication with the instrument.

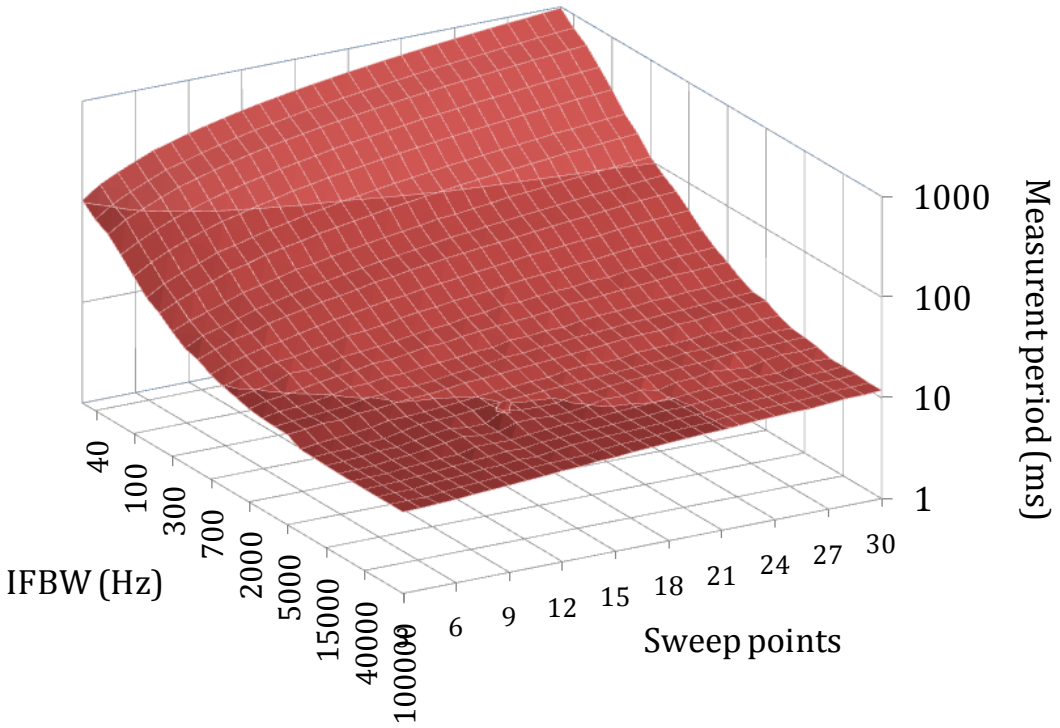



*Figure 4.30: Optimised VNA measurement period for various IF bandwidths and number of sweep points.*

Whilst the above graph would seem to suggest that more points is preferable to lower IFBW from a measurement time point of view, the repeatability data show that the number of points has almost no effect on the measurement precision. In fact, the optimum number of points appears to be reached by seven. After this, the measurement repeatability actually gets slightly worse. However, the repeatability is roughly proportional to the square root of the IFBW in both resonant frequency

and bandwidth. Therefore, the lowest IFBW should be used that still allows the desired measurement rate. The smallest possible repeatability with this VNA and resonator seems to be around 20 Hz. Considering that the resonant frequency of this sensor was 3.48 GHz, this represents a repeatability of 5 parts per billion (ppb) in resonant frequency and about 70 parts per million (ppm) in bandwidth for this resonator and VNA readout combination. This is extremely stable, indicating that MRS systems like this would be very useful for low noise sensing and precision control applications where absolute accuracy is not as important as noise. If the MRS system could be made cheaply enough, this advantage alone could give the technique mileage over traditional sensing methods in industrial control applications.

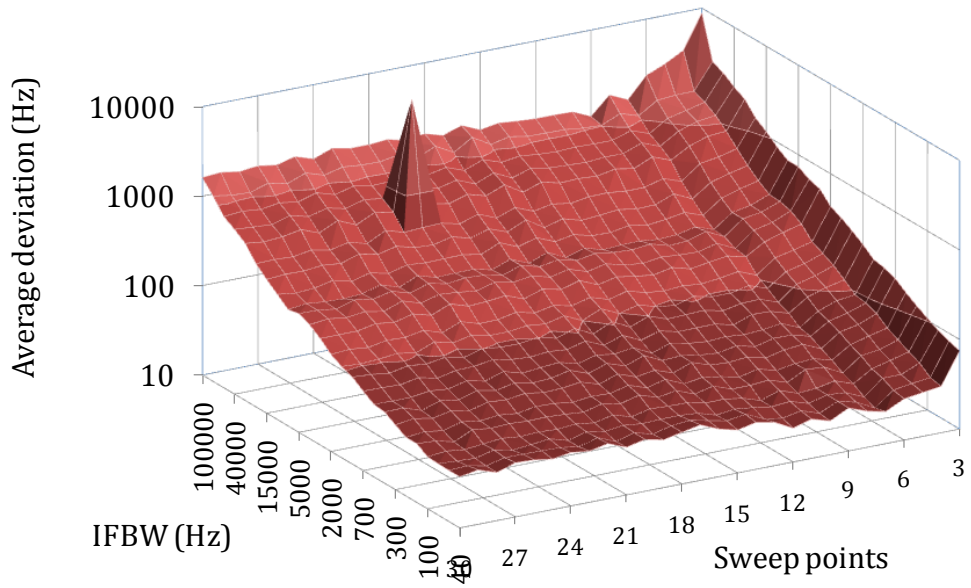



*Figure 4.31: Average point-to-point standard deviation in estimated resonant frequency values over 1000 consecutive samples. The peak is an outlier, but considering the 1000 averages might be caused by the presence of spurious signals in the IF receiver at a specific sweep frequency under those settings.*

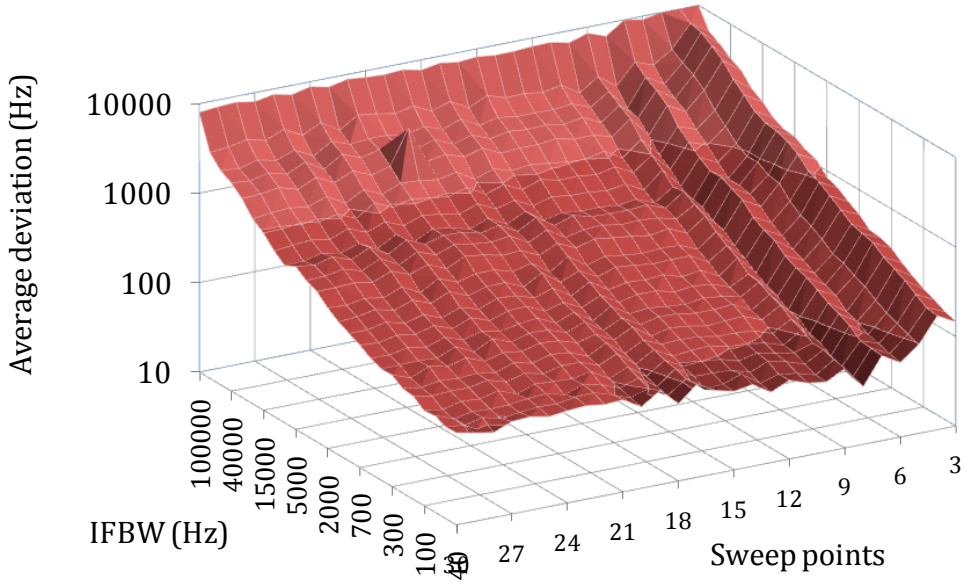



*Figure 4.32: Average absolute point-to-point standard deviation in estimated bandwidth values over 1000 consecutive samples.*

Repeatability is not the only aspect of measurement, however. If the estimated values are compared, it is apparent that there is some bias in the measurement with the number of sweep points (although not with IFBW). This is likely due to the simple but imperfect Lorentzian fitting algorithm, as discussed in the Software section. As the number of points is increased, estimates of bandwidth and resonant frequency seem to converge, assumedly to their 'true' values. It can be concluded that a reasonably large number of points is needed to ensure bandwidth estimates are unbiased (*i.e.,* accurate). With just a few points, this bias can be as high as 25% in bandwidth and 0.05% in resonant frequency, orders of magnitude higher than the precision. Note that the span used was the same in each case, so this bias is caused by sweep frequency resolution rather than coverage of the resonator spectral response.

Whilst this bias is probably constant (and would be cancelled in a differential perturbation measurement), this assumption is not definitive and cannot really be tested, as a perfectly repeatable perturbation does not exist. It is therefore wise to

mitigate the worst of the bias by over-specifying the number of points, at the expense of measurement rate.

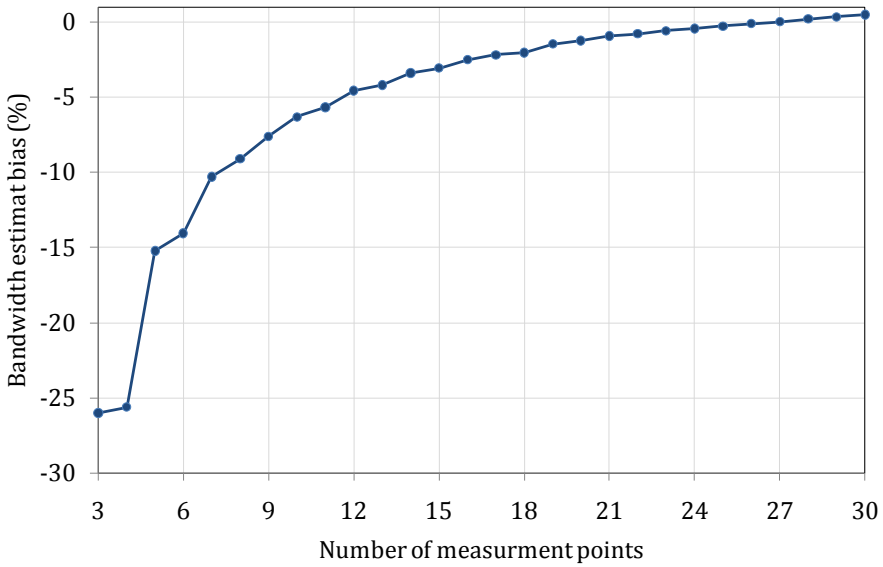



*Figure 4.33: Apparent measurement bias in bandwidth measurements.*

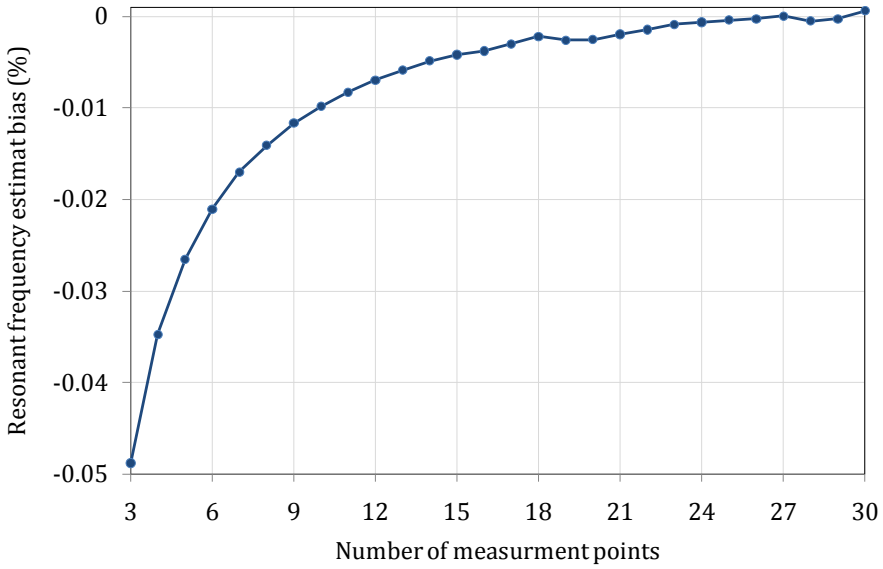



*Figure 4.34: Apparent measurement bias in resonant frequency.*

To summarise this section, software and techniques were developed to make VNA-based comparison measurements of MRS resonant frequency and bandwidth, optimised for speed, precision, and accuracy. External triggering mitigates sampling jitter and software pipelining reduces processing and transfer overheads.

IFBW was found to be crucial to measurement performance and should be minimised. Furthermore, the number of points should be carefully chosen to balance measurement rate against potential bias.

## 4.7 Static system noise comparison

To compare the measurement precision of the multitone system to the optimised VNA measurement, an experiment was designed to determine the levels of measurement noise present on a nominally static resonator. The experiment revealed some unexpected facts about resonator measurements in general, showing that the major source of noise in MRS systems below about 10 Hz appears to come from thermal fluctuations rather than the readout system – either VNA or multitone. The experiment further verified that the multitone system has comparable static performance to a VNA. The experiment also led to an interesting possibility. The precision of MRS measurements could be improved through perturbation signal modulation and 'lock-in' techniques. This could potentially make temporal MRS measurements useful for metrological as well as existing dynamic applications. This could be achieved by 'chopping' a sample perturbation on and off *(e.g.,* by mechanical removal and replacement).

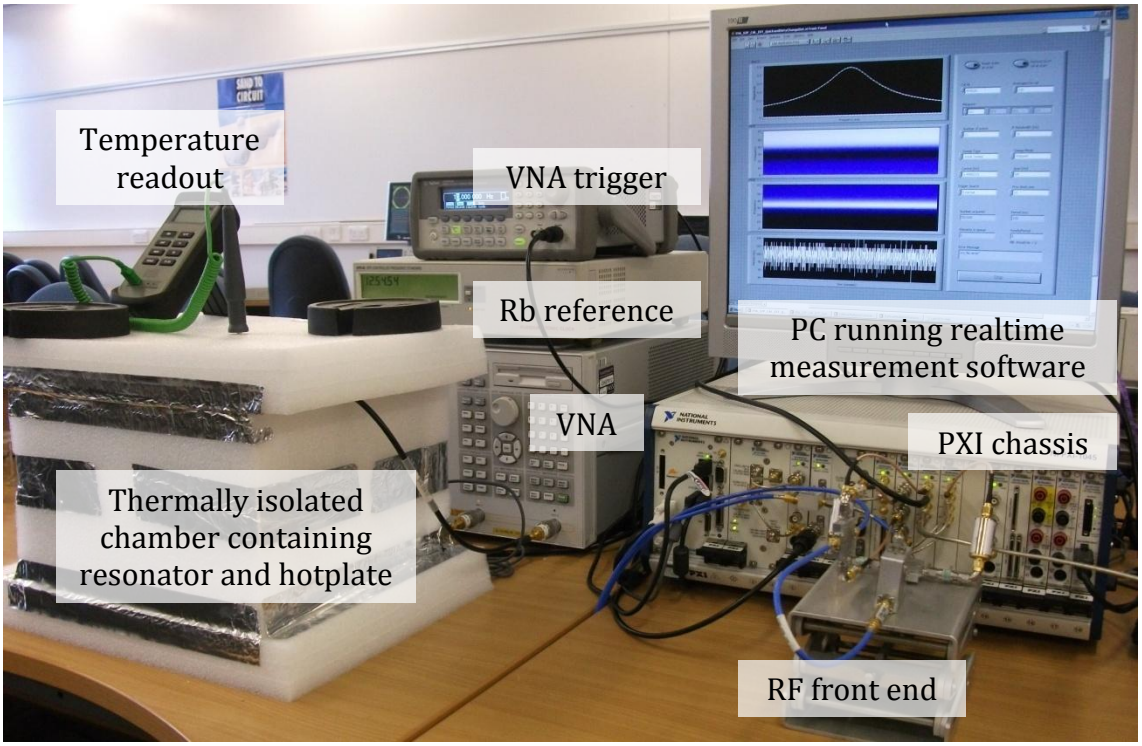



*Figure 4.35: Experimental setup for the VNA and experimental multitone system static testing experiments.*

An existing compact cylindrical resonator was measured at its 3.48 GHz $TM_{010}$ mode. The resonator was mechanically very robust and was precision machined from copper. The thick walls and high thermal conductivity of copper ensures that temperature fluctuations should cause minimal thermal expansion and contraction. There were no dielectric materials in the cavity, so the effect of thermal expansion on resonant frequency could be easily predicted. The resonator was placed on a warmplate (Minitube HT50) with thermostatic control set to 40°C. The hotplate and resonator were surrounded by a 5 cm thick insulating foam layer and the joins were sealed with aluminium tape.

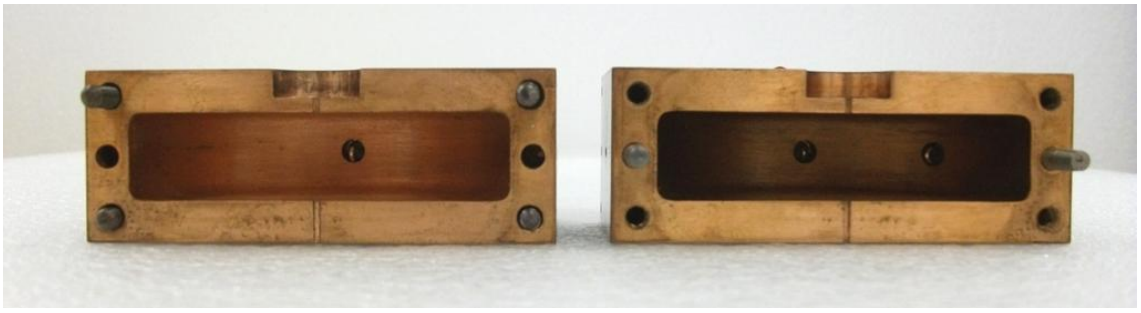


*Figure 4.36: Photograph of the two halves of the cylindrical resonator used for these experiments. This method of construction ensures that the radial-longitudinal $TM_{010}$ mode currents are not forced cross the join – preserving a high conductor Q.*

Phase noise in the local oscillator of any MRS readout method is theoretically impossible to separate from a 'real signal'. Furthermore, if locked to the same absolute frequency reference, two different MRS readout systems for estimating resonant frequency should read the same if there is no measurement bias. Therefore, in order to ascertain better the source of the measured noise and compare the accuracy of both systems, a 10 MHz Rubidium atomic clock reference (GPS-89, Spectracom, NY, USA) with global position system (GPS) lock was obtained. This was used as an alternative reference to which the PLL of the local oscillators in both systems could be locked.

Rubidium frequency references lock to a known hyperfine transition at 6.834 682 610 904 324 Hz [44]. Optically pumped electrons of a $^{87}$Rb plasma undergo a hyperfine transition when excited with microwaves at this frequency, causing a small change in the optical absorbency of the Rb vapour that can be detected using a photo cell. This shift in absorbance is used to tune a microwave synthesiser, which is then divided down to produce the 10 MHz standard used to discipline other oscillators. Rb oscillators are one of the cheapest precision timing sources available, and the technology has been miniaturised quite successfully – Symmetricom manufacture a miniature Rb oscillator just 40 x 35 x 11 mm, weighing only 35 g and consuming less than 120 mW [45]. When combined with the 1 pulse-per-second GPS signal derived from the orbiting Caesium atomic clock ensemble, frequency stability can be guaranteed be $< 1 \times 10^{-12}$ parts over periods

ranging from one minute to days and $< 3 \times 10^{-11}$ parts over periods from one second to one minute.

The VNA was configured as described above using an external trigger and LabVIEW realtime control, however raw spectral data were collected rather than being converted to resonant frequency on-line. This was to investigate the noise on both the raw scattering parameter estimate and the resonant frequency separately. A full 2-port short, load, open, through (SLOT) calibration was performed prior to measurements, extending the reference planes to the resonator coupling structures (made from bulkhead SMA connectors terminated in magnetic loops). As this was a long-term stability study, the measurement sample rate was set to 10 Hz and the IFBW 10 KHz. While not the fastest possible sample rate, it was chosen to demonstrate a good compromise between precision and speed and to avoid generating too much data over the run (even so, the uncompressed measurement file was over 1 GB). Over half a million samples were collected for each experiment, an average run time of about 15 hours.

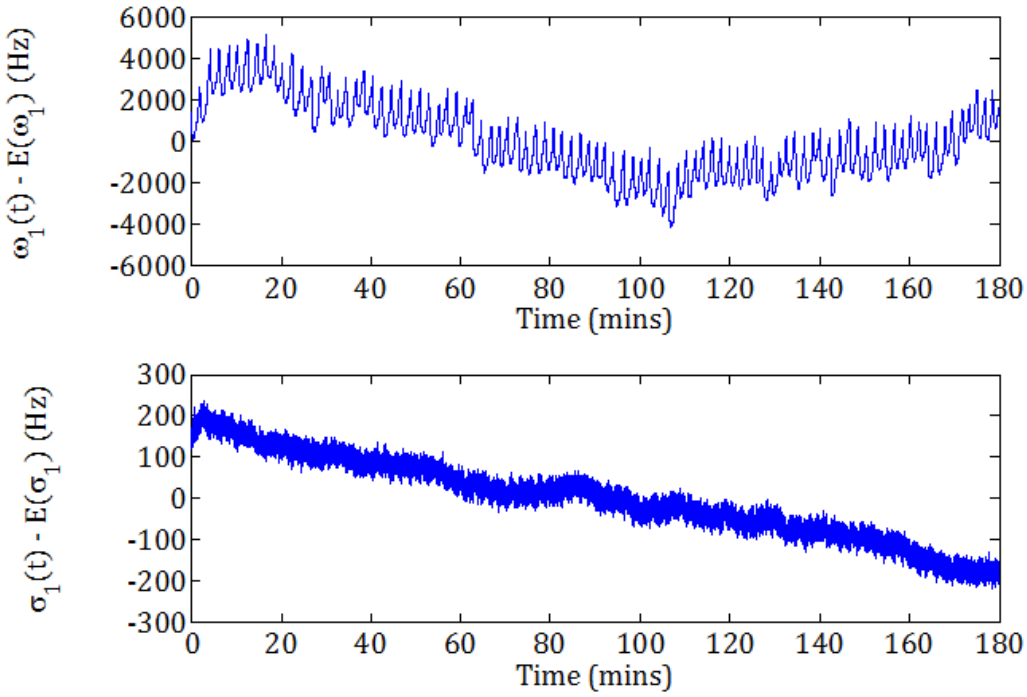



*Figure 4.37: Example of the noise measured in resonant frequency (top) and bandwidth (bottom) over a 3-hour period using the VNA.*

A multitone readout system based on the second hardware implementation was set up for narrowband measurement using digital upconversion around $f_s/4$. A 32 tone, 256-sample excitation was used, giving a spectral coverage of 12.5 MHz centred on resonance. As the insertion loss of the resonator was quite low, a single RF amplifier was used at the transmitter side to ensure the baseband signal amplitude made efficient use of the ADC code range, maximising the signal to quantisation noise ratio. Since the resonator was assumed stationary, a simple FFT-based spectral estimation algorithm (discussed in the Software chapter) was used to estimate the power spectral response from 256 successive averages to which a Lorentzian function was fit. This allowed the system to acquire the data, compute the spectrum and perform the curve fitting within a 20 ms time period whilst still guaranteeing real-time operation with no missed samples due to the fetch bottleneck. However, the number of samples collected per second with the multitone system is $256 \times 256 \times 50 \approx 3.3$ MS/s, whereas the maximum that could be collected is clearly 100 MS/s if all the data could be processed. This represents an information utilisation of just 3 %.

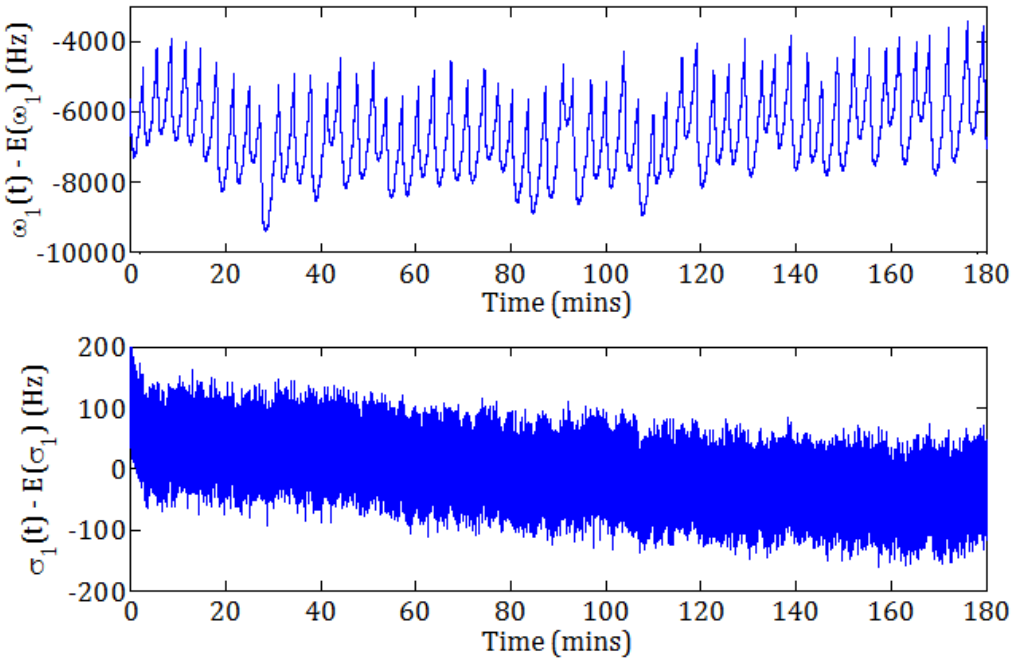



*Figure 4.38: Example of the noise measured in resonant frequency (top) and bandwidth (bottom) over a 3-hour period using the multitone system.*

Although these measurements were taken at different times, the experimental setup used for each readout system was identical. The same signal in resonant frequency is present in two independent measurement systems, regardless of whether they were locked to the Rb oscillator, so it could not be an artefact of the measurement system. Although the bandwidth noise on appears to be worse with the multitone system, it also samples 5 times quicker – noise power spectral density (PSD) was thus used to make a sample rate independent comparison.

To post process the VNA data, 1000 samples randomly selected from the whole dataset of spectral data were used to fit an 8-term complex-domain resonator model (this is elaborated on in later experiments). The parameters from this model were used to derive estimates for the resonator constants $K_0$ and $K_1$, which were applied to correct the rest of the spectral data before doing a simple 3-term complex-domain fit. This eliminated the effects of coupling reactance and the electrical length of the coupling structures in order to provide a very accurate absolute reference by which to compare the multitone system measurements.

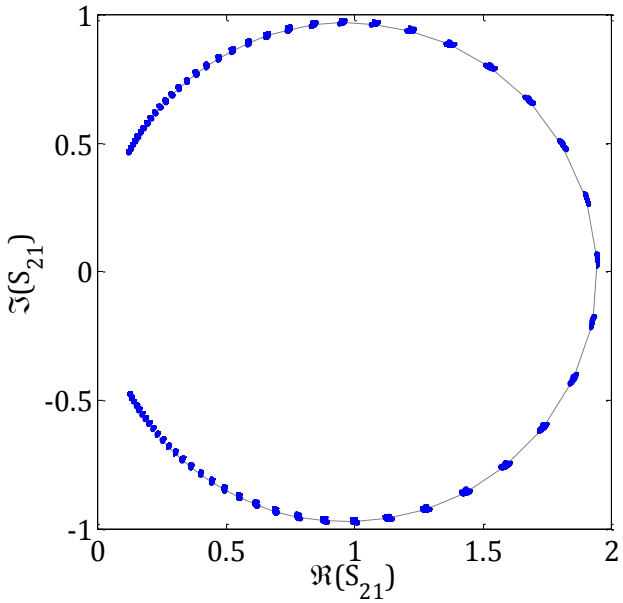



*Figure 4.39: Example of $K_0$ and $K_1$ normalised data over 1000 randomly selected data points, with 3-term fit shown in grey. The correlation between real and imaginary scattering parameter is evident from the scatter of individual points (blue).*

In order to investigate the nature of the measurement noise, power spectral density (PSD) estimates were made for the deviations in both resonant frequency and bandwidth (the relative errors):

$$\eta_\omega(t) = \frac{\omega_1(t) - E(\omega_1)}{E(\omega_1)}, \qquad \eta_\sigma(t) = \frac{\sigma_1(t) - E(\sigma_1)}{E(\sigma_1)} \qquad 4.10$$

The sample mean was used to estimate the expected value in each case. Thomson's multitaper method [46] was applied to DFT estimates over a logarithmic frequency grid derived using the Goertzel algorithm [47]. This method, which uses optimal frequency-domain tapers based on the orthogonal Slepian sequences, is a highly optimal, low bias spectral estimator for stationary ergodic random processes, an appropriate assumption in this case. The logarithmic frequency spacing is convenient for visualising noise behaviour over a range of time scales.

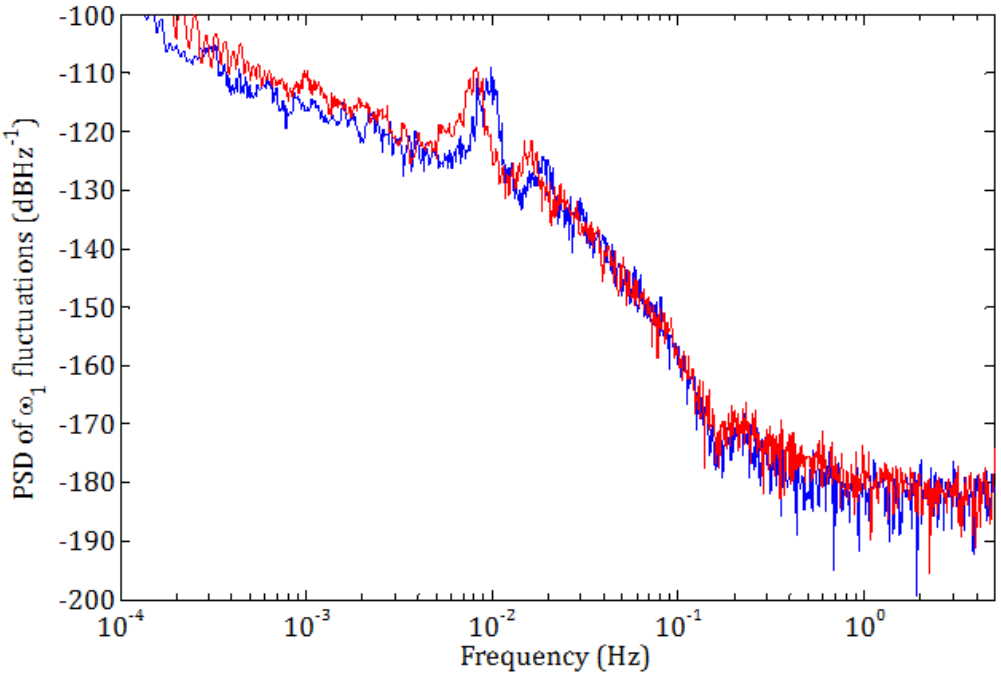



*Figure 4.40: PSD of resonant frequency fluctuations measured with a VNA. The blue curve is with the Rb oscillator lock, red internally referenced.*

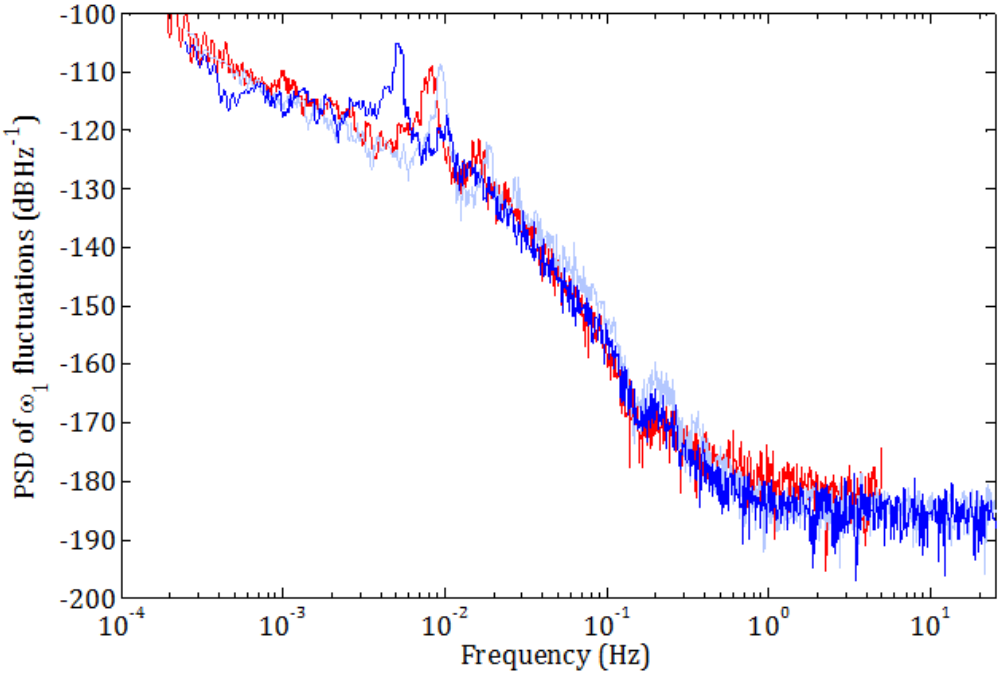



*Figure 4.41: PSD of resonant frequency fluctuations measured with the multitone system, the light blue curve is without Rb oscillator lock, dark blue with. Red is the VNA measured PSD for reference.*

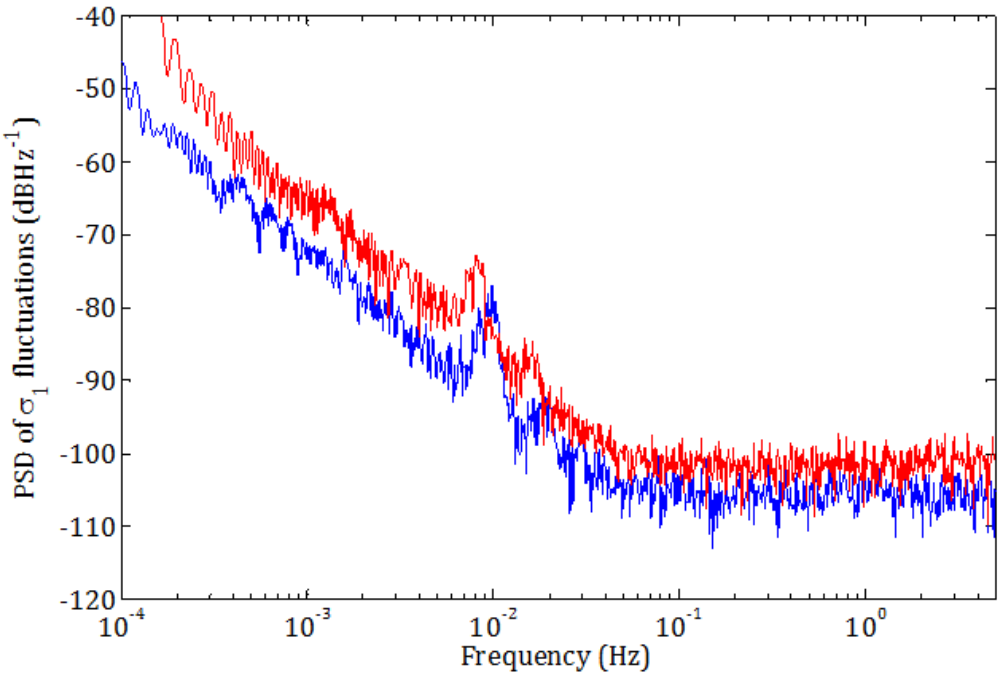



*Figure 4.42: PSD of bandwidth fluctuations measured with a VNA. The blue curve is with the Rb oscillator lock, red internally referenced.*

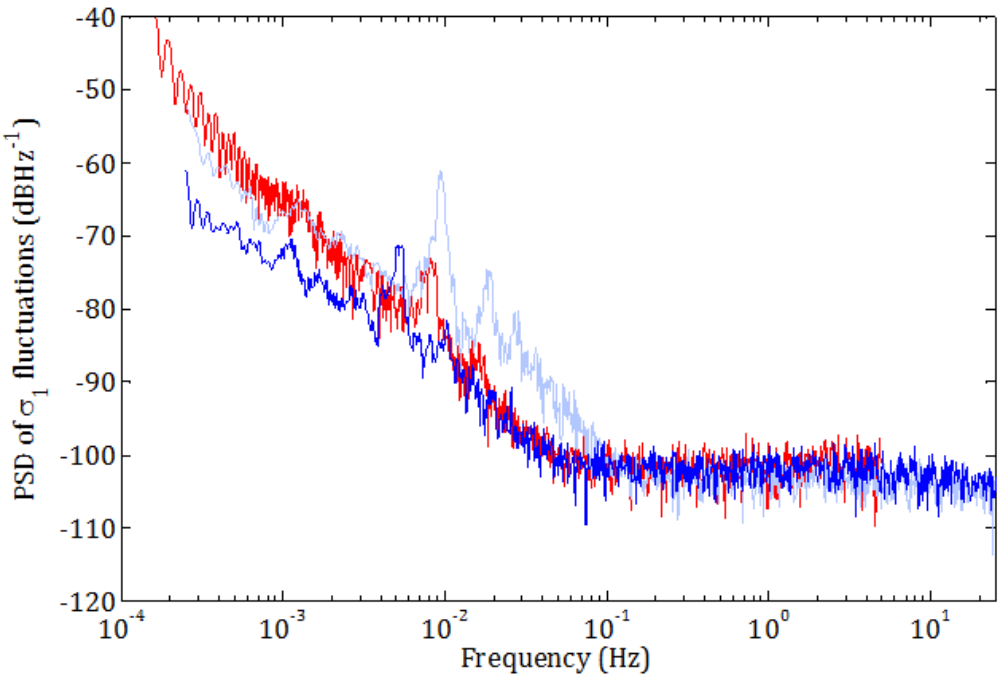



*Figure 4.43: PSD of bandwidth fluctuations measured with the multitone system, the light blue curve is without Rb oscillator lock, dark blue with. Red is the VNA measured PSD for reference.*

The power spectral density analysis is intriguing. The $\omega_1$ noise appears to be relatively flat, and extremely low (one part per billion corresponds to a standard error of $-180$ dB) on timescales shorter than about 5 seconds (0.2 Hz). Between about 0.01 Hz and 0.1 Hz the spectral slope is approximately $-20$ dB/dec. This would be consistent with close-in local oscillator phase noise or filtering by a long time-constant first order system. At very low frequencies the spectrum tends to a $1/f$ distribution, or $-10$ dB/dec. Furthermore, the results with and without the Rb oscillator imply that low frequency oscillator phase noise does not cause the resonant frequency fluctuations; there is no statistically significant difference between the power spectra measured with and without it. Some improvement is seen in bandwidth measurements, however, probably due to the improved noise floor brought about by the better short-term phase noise of the Rb oscillator reference. This suggests that the only significant effect of phase noise is its conversion to uncorrelated amplitude noise in the spectral measurements, rather than causing apparent shifts in resonant frequency.

In comparison, the PSDs of the VNA and multitone system are very similar – showing that the multitone readout precision is comparable to the VNA throughout the input frequency range. The multitone system PSD extends to a higher frequency because it has a higher sampling rate (50 Hz as opposed to 10 Hz). There is a small discrepancy in $\omega_1$ noise at about 0.2 Hz, however, this is on the scale of the 95 % confidence intervals for these spectra (around $+4/-3$ dB). There is a peak difference of about 10 dB in bandwidth noise, however, in the 0.01-0.1 Hz range when the multitone system is not locked to the Rb oscillator. The fact that this coupling appears to go away when the Rb oscillator is used is curious. Judging by the increased signal seen in this region as well, the cause of this can be put down to noise coupling between frequency and bandwidth, *i.e.,* some of the resonant frequency fluctuations are being coupled to the bandwidth dimension. Although some natural correlation is expected, the discrepancy suggests that, in the multitone system, this may be caused by coupling between the *I* and *Q* channels – *i.e.,* AM to PM conversion. The most significant causes of this effect are quadrature mismatch and nonlinearity. Hence, the poorer linearity of the

multitone system is causing more resonant frequency fluctuations to be converted to bandwidth fluctuations than in the VNA. Although the effect is small in magnitude, it could be prevented by correcting (or predistorting for) quadrature errors and nonlinearity (a topic covered in more detail in the Chapter 5).

In order to explore the nature of the noise source further, more statistical information was computed. The Pearson correlation coefficient [48] is a measure of linear relationship between two random variables. For measuring the correlation between two vectors, such as the vector of scattering parameter estimates against frequency, the Pearson correlation matrix can be derived by scaling by the diagonal of the individual vector covariance matrices, $\Sigma_{xx}$:

$$\mathbf{r}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(\mathbf{x}_i - \bar{\mathbf{x}})}{diag(\Sigma_{xx})}\frac{(\mathbf{y}_i - \bar{\mathbf{y}})^T}{diag(\Sigma_{yy})}, \qquad 4.11$$

this gives a matrix where the diagonal is unity and the off-diagonal elements vary between $-1$ (perfect negative correlation) and $+1$ (perfect positive correlation.

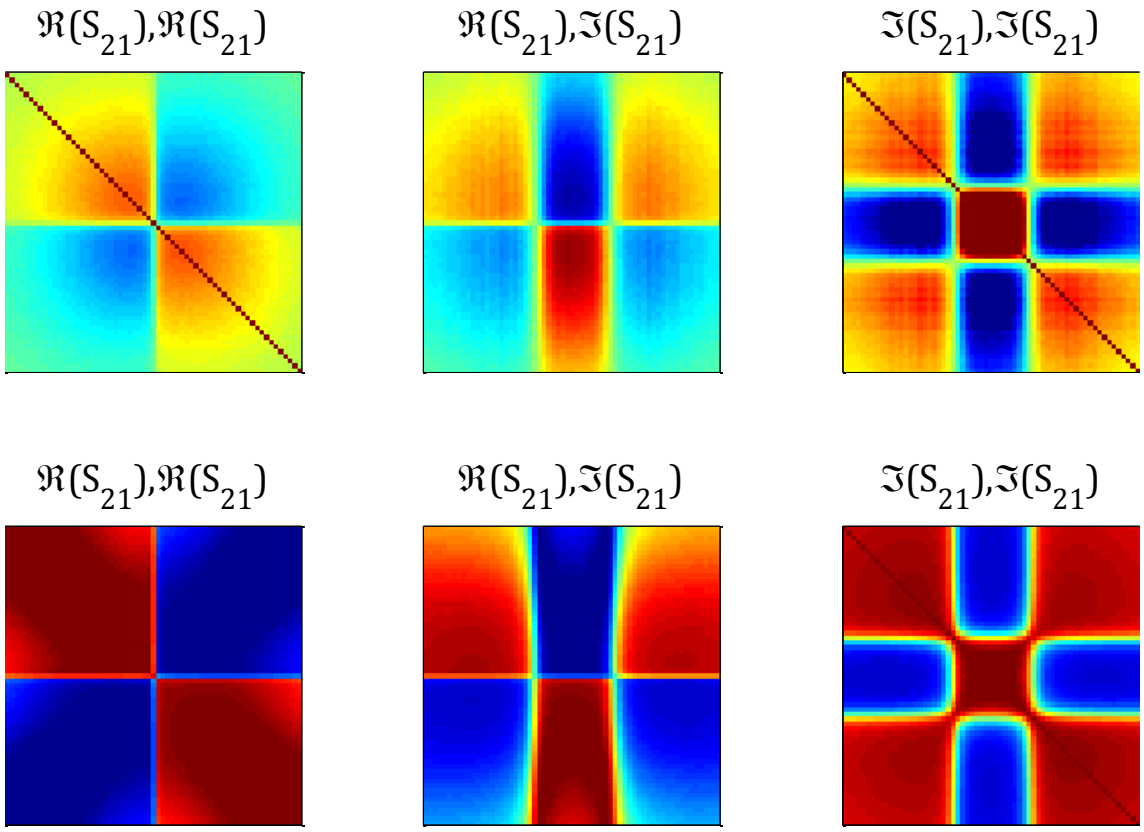



*Figure 4.44: Pearson correlation matrices over the measured spectrum for linearly de-trended VNA data over 30 seconds (top row), and 30 minutes (bottom row). Dark blue represents perfect negative correlation, $r = -1$, and dark red being perfect positive correlation, $r = 1$. On the left of the row is the correlation between frequencies in the real part of the spectral response, in the middle is the correlation between frequencies and between real and imaginary, and on the right is the correlation between frequency in the imaginary part of the spectral response.*

Each set of corresponding pixels over the three images in a row corresponds with a 2×2 covariance matrix in the real and imaginary scattering parameter dimensions:

$$\Sigma_{R_i I_j} = \begin{bmatrix} \sigma_{R_i R_j} & \sigma_{R_i I_j} \\ \sigma_{I_i R_j} & \sigma_{I_i I_j} \end{bmatrix} \quad \forall\, i,j = 1..m, \tag{4.12}$$

$m$ being the number of frequency sweep points and $R$ and $I$ denoting $\Re(S_{21})$ and $\Im(S_{21})$, respectively. These many covariance matrices each have an equivalent 'uncertainty ellipse' in the complex plane. This was touched upon in the Applications chapter when looking at previous work on kinetic inductance detectors. The uncertainty ellipse is an accepted way of visualising the correlation between measured noise sources in these resonant sensors. The above approach, with three separate Pearson matrices, generalises this analytic approach to multifrequency measurement (which kinetic inductance detector readouts do not use, in general, hence this method has not been considered previously).

According to this analysis, there is a very high degree of correlation between adjacent frequency points on the resonance curve, and that correlation increases over longer periods. The change in sign of the correlation in $\Re(S_{21})$ marks the resonant frequency, and two changes in correlation sign in $\Im(S_{21})$ mark the 3dB bandwidth. Even on the 30-second time scale, there is a very high correlation between adjacent imaginary parts of the scattering parameter within the resonator bandwidth. This is consistent with small changes in resonant frequency, since the change in $\Im(S_{21})$ is proportional to a change in resonant frequency to first order, as shown in the Theory section by Taylor series expansion. Over large time scales, there is generally a very high correlation over the entire spectral response. The structure of these correlations confirms that the fluctuations measured in resonant frequency are not due to noise in the receiver. This would not cause correlations between frequency points, even if it might cause correlations between the real and imaginary scattering parameter estimates. This further supports the hypothesis that, on a time scale of minutes and above, the measurement noise is dominated by small variations in resonant frequency arising naturally in the resonant sensor itself.

There is a clear signal with a period around 1.5 minutes, visible on the resonant frequency vs. time trace and on the noise power spectral density graph. Some variation from measurement to measurement is seen. After eliminating the measurement systems as an explanation, this had to be due to the warmplate thermostatic loop. The triangle-wave like heating and cooling cycle caused by its

hysteretic switching control method. As this control loop was specified to keep the temperature nominally within ± 0.1°C, this peak-to-peak variation of about 3 KHz represents the resonant frequency change corresponding to around a 0.1 K temperature variation. Temperature changes cause cavity resonators to expand and contract, changing their electrical dimensions and therefore resonant frequency. The thermal expansion coefficient of copper at 40°C is $\beta_L \approx 17$ ppmK$^{-1}$ [49]. The resonant frequency and bandwidth for this air-filled $TM_{010}$ mode resonator are given by the following expressions, derived from the analytical solutions of the Helmholtz wave equation in cylindrical coordinates: [18]

$$\omega_0 = \frac{c\alpha}{a}, \quad \sigma_0 = \frac{R_s}{\mu_0}\frac{c^2\alpha'^2}{a^3\omega^2} = \sqrt{\frac{2c}{a^3\mu_0\sigma_{Cu}}}\left(\frac{\alpha'^2}{2\alpha^{\frac{3}{2}}}\right), \qquad \left(\omega = \omega_0,\ R_s = \sqrt{\frac{\omega_0\mu}{2\sigma_{Cu}}}\right) \qquad 4.13$$

where $a \approx 33$ mm is the cavity radius, $c$ is the speed of light, $\sigma_{Cu} \approx 5.6 \times 10^7\ S/m$ is the conductivity of the cavity walls, $\alpha' = 3.831\ldots$ and $\alpha = 2.404\ldots$ are constants, and $\mu_0 = 4\pi \times 10^{-7}\ H/m$ is the permeability of free space. Thus, both bandwidth and resonant frequency are inversely proportional to the cavity radius. Also, by the definition of thermal expansion, the change in radius for a given change in temperature, $da/dT = a\beta_L$. Theory thus predicts that the peak-to-peak deviations in resonant frequency are about $\sigma_{\omega_1} \approx 3.6$ KHz for a temperature variation of 0.1 K. This agrees quite closely with the observations – suggesting if anything that the true temperature variation over the heating and cooling cycle is slightly less than 0.1 K.

Based on this quite significant change with temperature, it is logical to estimate what the effect random temperature fluctuations might have on the measured resonant frequency and bandwidth. Using Equation 3.88, combined with the above expressions, the variance of the noise induced in resonant frequency and bandwidth due to thermal fluctuations in a $TM_{010}$ mode resonator is given by:

$$\frac{\sigma_{\omega_1}^2}{\sigma_T^2} = \frac{c^2\alpha^2}{a^2}\beta_L^2, \quad \frac{\sigma_{\sigma_1}^2}{\sigma_T^2} = \frac{c}{a^3\mu_0\sigma_{Cu}}\left(\frac{9\alpha'^4}{8\alpha^3}\right)\beta_L^2, \quad \frac{\sigma_{\omega_1\sigma_1}}{\sigma_T^2} = \sqrt{\frac{2c^3}{a^3\mu_0\sigma_{Cu}}}\left(\frac{3\alpha'^2}{4\sqrt{\alpha}}\right)\beta_L^2 \tag{4.14}$$

$$\sigma_{\omega_1}^2 \approx 1.7\times10^9\cdot\sigma_T^2, \qquad \sigma_{\sigma_1}^2 \approx 0.22\cdot\sigma_T^2, \qquad \sigma_{\omega_1\sigma_1} \approx 2.4\times10^4\cdot\sigma_T^2$$

where $\sigma_T^2$ is the variance of the random temperature fluctuations. Whilst temperature induced variance in bandwidth is predicted to be negligible, a significant correlation, $\sigma_{\omega_1\sigma_1}$, between bandwidth and resonant frequency is expected due to temperature fluctuation.

Based on the magnitude of the thermal noise predicted above, combined with the $-20$ dB/dec trend in the noise spectral density on the minute time scale, it implies that the source of the noise is random temperature fluctuations that are filtered by a long thermal time constant – *i.e.,* the combined the thermal mass of the resonator and metal hotplate. The filtering effect comes from applying Newton's law of cooling by convection:

$$\frac{dT(t)}{dt} = -\frac{hA}{mc_p}\Delta T(t), \tag{4.15}$$

which states that the rate of change of temperature with time is proportional to the convective heat transfer coefficient $h$, multiplied by the body's surface area $A$ divided by the body's specific heat capacity $c_p$ times its mass $m$ (the body being the combined resonator-hotplate thermal system in this case). $\Delta T(t)$ is the thermal gradient between the body and the environment. Thus if the environmental temperature is considered to be random white noise, and the system time constant $\tau = mc_p/hA$ is large, it will act as a first order low pass system with a low cut-off frequency; the roll-off from which is $-20$ dB/dec.

This provides an explanation for the measured noise in resonant frequency, and gives weight to the conclusion that the flat, white noise section represents the measurement system noise floor whereas the $1/f^2$ noise is generated by thermal fluctuations in the resonant sensor. This demonstrates that the new MRS measurement system is intrinsically capable of resolving part-per-billion changes

in resonant frequency – about 180 dB or 30 bits of dynamic range – at 1 Hz rate. Since the measurement noise appears to be white, this figure should decrease by 10 dB for each 10 × in sampling rate. Furthermore, if the multitone system was able to achieve 100 % sample utilisation instead of the 3 % currently possible, this figure could theoretically be improved by 15 dB, assuming the white noise distribution continues at higher frequencies. Few sensors and sensor readout systems can boast this intrinsic level of precision; as it stands this is $40 - 60$ dB more dynamic range than cutting-edge optical image sensors, for example [50], [51]. This is why the MRS technique is desirable as a means to make astronomical detectors using microwave kinetic inductance detector arrays.

Converting this precision to that of the desired measurand estimate is evidently application-specific; however, based on a simple linear conversion, and assuming that the maximum possible signal is limited to the 100 MHz bandwidth of the multitone system, this translates to a dynamic range, *e.g.,* for real permittivity, of 150 dB – around 30 parts-per-billion (ppb), or 25 bits. If the noise statistics for bandwidth are assumed to be constant, and the largest bandwidth practically measurable is around 50 MHz (a factor of 100 more than this sensor has when empty), this implies that the dynamic range for bandwidth perturbation (*e.g.,* imaginary permittivity) is around 140 dB – around 100 ppb, or 23 bits. However, when the bandwidth is time-varying, the situation is complicated by the fact that the SNR is proportional to the insertion loss and therefore (approximately) inversely proportional to bandwidth, a situation known as heteroscedasticity (amplitude-dependent statistics). Thus, at the largest bandwidth practically measurable this would result in 40 dB degradation in the noise floor and hence dynamic range. Thus at maximum bandwidth perturbation the resonant frequency perturbation dynamic range could be reduced to 110 dB, and 100 dB for bandwidth perturbation. Fortunately, the noise present on the multitone system is limited by the input ADC quantisation noise. This situation could therefore be alleviated to some degree by using a baseband variable gain amplifier (VGA).

The above discussion ignores thermal noise, so for longer measurements the resonant sensor would need to be perfectly thermally stabilised throughout a

measurement to achieve the same precision. However, such thermal stability is difficult to achieve in practice, especially over long periods. This raises an interesting possibility. If the main source of noise is thermal fluctuations, and it is possible to control temperature variations adequately enough to limit this noise to low frequencies, it makes a lot of sense to modulate the input to the sensor at a frequency above this cut-off point. Judging by these data and the level of thermal fluctuations of this resonator, such a frequency would be about 0.1-1 Hz in this case. This is a key insight to arise out of this work.

Whist it is difficult to generalise, it is likely that most resonant sensors without specific sensitivity to known sources of noise are dominated by thermal expansion-contraction noise, making this conclusion of general significance. Examples of other intrinsic sources include microphonics (vibration) in the case of microwave microscope sensors, or two level system fluctuations at the substrate-metal interface in the case of MKIDs. This also implies that, without good reason to do otherwise, simple and robust metal cavity resonators should be preferred to more complex sensor designs from a precision point of view.

In smaller resonators with a correspondingly shorter thermal time constant, or in less well controlled environments having higher thermal fluctuation, this ‘knee’ in the noise PSD would be shifted to higher frequencies, making higher speed sampling a necessity. This is an unexpected application of the fast-sampling multitone system. Faster sampling could possibly *increase* the dynamic range of nominally ‘static’ metrological measurements, making the multitone system appealing in a general laboratory setting as well as in specific high-speed sensing applications.

## 4.8 Chapter references


[1] R. I. Lackey and D. W. Upmal, "Speakeasy: the military software radio," *IEEE Communications Magazine*, vol. 33, no. 5, pp. 56-61, May 1995.

[2] P. B. Kenington, *RF and baseband techniques for software defined radio*. Artech House, 2005.

[3] W. H. . Tuttlebee, "Software-defined radio: facets of a developing technology," *IEEE Personal Communications*, vol. 6, no. 2, pp. 38-44, Apr. 1999.

[4] "Test and Measurement Products | Tektronix." [Online]. Available: http://www.tek.com/home/products.html. [Accessed: 07-Sep-2011].

[5] "Oscilloscope - LeCroy." [Online]. Available: http://www.lecroy.com/oscilloscope/. [Accessed: 07-Sep-2011].

[6] C. F. Coombs, *Electronic instrument handbook*. McGraw-Hill, 2000.

[7] A. L. Cullen and T. Y. An, "MICROWAVE CHARACTERISTICS OF THE SCHOTTKY-BARRIER DIODE POWER SENSOR.," *IEE Proceedings H: Microwaves Optics and Antennas*, vol. 129, no. 4, pp. 191-198, 1982.

[8] M. B. Ewing and D. D. Royal, "A highly stable cylindrical microwave cavity resonator for the measurement of the relative permittivities of gases," *The Journal of Chemical Thermodynamics*, vol. 34, no. 7, pp. 1073-1088, Jul. 2002.

[9] G. F. Engen, "A (historical) review of the six-port measurement technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 45, no. 12, pp. 2414-2417, Dec. 1997.

[10] K. Haddadi, M. M. Wang, D. Glay, and T. Lasri, "Ultra wide-band four-port reflectometer using only two quadratic detectors," in *Microwave Symposium Digest, 2008 IEEE MTT-S International*, 2008, pp. 379-382.

[11] G. Wei, J.-D. Xu, C.-Y. Wu, and J.-X. Yang, "Designing six-port measurement system for measuring electrical thickness of radome and permittivity of materials," *Qiangjiguang Yu Lizishu/High Power Laser and Particle Beams*, vol. 19, no. 8, pp. 1347-1351, 2007.

[12] X.-P. Hu, "USING SIX-PORT REFLECTOMETER MEASUREMENT OF COMPLEX DIELECTRIC CONSTANT.," *IEEE Transactions on Instrumentation and Measurement*, vol. 36, no. 2, pp. 537-539, 1986.

[13] U. Stumper, "Six-port and four-port reflectometers for complex permittivitymeasurements at submillimeter wavelengths," *IEEE Transactions on Microwave Theory and Techniques*, vol. 37, no. 1, pp. 222-230, Jan. 1989.

[14] O. Benzaim, K. Haddadi, M. M. Wang, M. Maazi, D. Glay, and T. Lasri, "Scanning Near-Field Millimeter-Wave Microscope: Application to a Vector-Coding Technique," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 11, pp. 2392-2397, Nov. 2008.

[15] M. Maazi, O. Benzaim, D. Glay, and T. Lasri, "Detection and characterization of buried macroscopic cracks inside dielectric materials by microwave techniques and artificial neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 12, pp. 2819-2826, 2008.

[16] J. R. Juroshek and C. A. Hoer, “A Dual Six-Port Network Analyzer Using Diode Detectors,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 32, no. 1, pp. 78- 82, Jan. 1984.

[17] C. Li, H. Zhang, and P. Wang, “A novel six-port circuit based on four quadrature hybrids,” *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 20, no. 1, pp. 128-132, Jan. 2010.

[18] D. M. Pozar, *Microwave engineering*. J. Wiley, 2005.

[19] F. Harris, “Digital filter equalization of analog gain and phase mismatch inI-Q receivers,” in *, 1996 5th IEEE International Conference on Universal Personal Communications, 1996. Record*, 1996, vol. 2, pp. 793-796 vol.2.

[20] J. Tubbax, B. Come, L. Van der Perre, S. Donnay, M. Engles, and C. Desset, “Joint compensation of IQ imbalance and phase noise,” in *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual*, 2003, vol. 3, pp. 1605- 1609 vol.3.

[21] Lei Sun, Zhijun Li, and Miao Yang, “Compensation of IQ Imbalance Based on a Simplified Blind Source Separation Method,” in *5th International Conference on Wireless Communications, Networking and Mobile Computing, 2009. WiCom ’09*, 2009, pp. 1-5.

[22] A. R. Wright and P. A. Naylor, “Blind IQ mismatch compensation in OFDM direct conversion receivers,” in *DSPenabledRadio, 2005. The 2nd IEE/EURASIP Conference on (Ref. No. 2005/11086)*, 2005.

[23] S. Doyle, P. Mauskopf, J. Naylon, A. Porch, and C. Duncombe, “Lumped Element Kinetic Inductance Detectors,” *Journal of Low Temperature Physics*, vol. 151, pp. 530-536, Jan. 2008.

[24] “AD7760 | 2.5 MSPS, 24-Bit, 100 dB Sigma-Delta ADC with On-Chip Buffer | All A/D Converters | Analog to Digital Converters | Analog Devices.” [Online]. Available: http://www.analog.com/en/analog-to-digital-converters/ad-converters/ad7760/products/product.html. [Accessed: 06-Oct-2011].

[25] F. L. Walls, A. J. . Clements, C. M. Felton, M. A. Lombardi, and M. D. Vanek, “Extending the range and accuracy of phase noise measurements,” in *Frequency Control Symposium, 1988., Proceedings of the 42nd Annual*, 1988, pp. 432-441.

[26] Agilent Literature Number 5965-7917E "Network Analyser Basics", 2004. Available: http://cp.literature.agilent.com/litweb/pdf/5965-7917E.pdf

[27] “Application Notes (Rohde & Schwarz International - Service & Support - Downloads).” [Online]. Available: http://www2.rohde-schwarz.com/en/service_and_support/Downloads/Application_Notes/. [Accessed: 06-Oct-2011].

[28] “E5071B ENA RF Network Analyzer [Discontinued] | Agilent.” [Online]. Available: http://www.home.agilent.com/agilent/product.jspx?nid=-536902663.536881709.00&lc=eng&cc=GB. [Accessed: 06-Oct-2011].

[29] Schlaerth et al., “A Millimeter and Submillimeter Kinetic Inductance Detector Camera,” *Journal of Low Temperature Physics*, Feb. 2008.

[30] Won Namgoong and T. H. Meng, “Direct-conversion RF receiver design,” *IEEE Transactions on Communications*, vol. 49, no. 3, pp. 518-529, Mar. 2001.

[31] “ADL5375 | 400 MHz to 6 GHz Broadband Quadrature Modulator | Modulators / Demodulators | RF / IF ICs | Analog Devices.” [Online]. Available: http://www.analog.com/en/rfif-

components/modulatorsdemodulators/adl5375/products/product.html. [Accessed: 06-Oct-2011].

[32] G. Avenier et al., "0.13 m SiGe BiCMOS Technology Fully Dedicated to mm-Wave Applications," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 9, pp. 2312-2321, Sep. 2009.

[33] "AD9957 | 1 GSPS Quadrature Digital Upconverter with 18-Bit IQ Data Path and 14-Bit DAC | Direct Digital Synthesis ( DDS) & Modulators | RF / IF ICs | Analog Devices." [Online]. Available: http://www.analog.com/en/rfif-components/direct-digital-synthesis-dds/ad9957/products/product.html. [Accessed: 06-Oct-2011].

[34] J. Naylon and A. Porch, "UK patent application number GB1109685.6: Apparatus and method for estimating a characteristic of a microwave resonant device," 2011.

[35] "National Instruments: Test, Measurement, and Embedded Systems." [Online]. Available: http://www.ni.com/. [Accessed: 06-Oct-2011].

[36] "PXI hardware and software specifications." [Online]. Available: http://www.pxisa.org/Specifications.html. [Accessed: 06-Oct-2011].

[37] "ADL5375 | 400 MHz to 6 GHz Broadband Quadrature Modulator | Modulators / Demodulators | RF / IF ICs | Analog Devices." [Online]. Available: http://www.analog.com/en/rfif-components/modulatorsdemodulators/adl5375/products/product.html. [Accessed: 06-Oct-2011].

[38] "ADL5380 | 400 to 6000 MHz Quadrature Demodulator | Modulators / Demodulators | RF / IF ICs | Analog Devices." [Online]. Available: http://www.analog.com/en/rfif-components/modulatorsdemodulators/adl5380/products/product.html. [Accessed: 06-Oct-2011].

[39] "Marki Microwave: Image-Reject / IQ." [Online]. Available: http://www.markimicrowave.com/3457/Image-Reject_/_IQ.aspx. [Accessed: 06-Oct-2011].

[40] "Owen Splitter - Microwave Encyclopedia - Microwaves101.com." [Online]. Available: http://www.microwaves101.com/encyclopedia/Resistive_splitter2.cfm. [Accessed: 06-Sep-2011].

[41] R. Inoue, K. Miwa, H. Kitano, A. Maeda, Y. Odate, and E. Tanabe, "Highly accurate and real-time determination of resonant characteristics: complex linear regression of the transmission coefficient," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 9, pp. 2163- 2168, Sep. 2004.

[42] P. M. Oliveira and V. Barroso, "Uncertainty in the time-frequency plane," in *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing, 2000*, 2000, pp. 607-611.

[43] Agilent Application Note 1363-1, "Understanding and Improving Network Analyser Dynamic Range", 2000. Available: http://cp.literature.agilent.com/litweb/pdf/5980-2778EN.pdf

[44] C. Mandache et al., "Comparison with an uncertainty of 2×10-16 between two primary frequency standards," in *Frequency Control Symposium and Exposition, 2005. Proceedings of the 2005 IEEE International*, 2005.

[45] "CSAC - SA.45s - Chip Scale Atomic Clock - Symmetricom." [Online]. Available: http://www.symmetricom.com/products/frequency-references/chip-scale-atomic-clock-csac/SA.45s-CSAC/. [Accessed: 06-Oct-2011].

[46] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055- 1096, Sep. 1982.

[47] G. Goertzel, "An Algorithm for the Evaluation of Finite Trigonometric Series," *The American Mathematical Monthly*, vol. 65, no. 1, pp. 34-35, Jan. 1958.

[48] S. J. Devlin, "Robust Estimation and Outlier Detection with Correlation Coefficients," *Biometrika*, vol. 62, no. 3, pp. 531–545, 1975.

[49] F. C. Nix and D. MacNair, "The Thermal Expansion of Pure Metals: Copper, Gold, Aluminum, Nickel, and Iron," *Physical Review*, vol. 60, no. 8, p. 597, Oct. 1941.

[50] A. Spivak, A. Belenky, A. Fish, and O. Yadid-Pecht, "Wide-Dynamic-Range CMOS Image Sensors—Comparative Performance Analysis," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2446-2461, Nov. 2009.

[51] T. Yamada, S. Kasuga, T. Murata, and Y. Kato, "A 140dB-Dynamic-Range MOS Image Sensor with In-Pixel Multiple-Exposure Synthesis," in *Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International*, 2008, pp. 50-594.

# 5 Software

Moving to an SDR architecture for MRS readout results in a higher flexibility and lower cost by exchanging analogue complexity in favour of DSP. The software algorithms that facilitate and enhance measurement performance of such a system are therefore crucial to the success of this approach. This chapter is thus devoted to the development and testing of the multitone measurement system software.

In the Theory section, the sensing process is represented as a multistage model, linking measurand to eigenvalue, eigenvalue to scattering parameter, scattering parameter to time-domain voltages. The software must unravel these processes, one by one, in reverse; this chapter is organised in this way, too. With hardware now capable of generating and receiving broadband, multifrequency signals, Section 5.1 looks at the task of estimating the spectral response from their time domain measurement. In order to keep the system simple, spectral estimation must rest on the assumption that the excitation is perfectly known. Thus, Section 5.2 ensures this is not a far-fetched assumption by looking at ways to use software to correct for the imperfections in the signal generation process.

Once a good spectral response is available, the next step is to turn it into an eigenvalue. Section 5.3 discusses how to go about this, from simple, tried-and-tested methods like curve fitting, through to novel, non-parametric methods suitable for the very fastest MRS readout rates and implementation on embedded processing platforms such as field-programmable gate arrays (FPGA).

## 5.1 Spectral response estimation

Without a sufficiently accurate and low noise estimate of the resonator spectral response, no amount of signal processing will be able to produce measurements with instrument quality precision. Therefore, the first and most important issue to be addressed in this chapter is how to exploit best the powerful broadband signal generation and measurement capability of the SDR system to produce an accurate resonator spectral estimate – ideally in a fraction of the time required by traditional methods.

As discussed extensively in the Theory chapter, the response of a microwave resonator to voltage wave excitation can be accurately modelled parametrically, including a number of real-world imperfections such as coupling reactance, electrical length of connecting lines and finite receiver directivity. This spectral response is well approximated as a linear system, as long as distortion in the transmitter and receiver is minimised. The task of estimating parameters for a linear system from excitation and response signals is known as linear system identification. The theory of linear system identification therefore underpins this section, so a short introduction will be given in Subsection 5.1.1.

Spectral response estimation is divided into three topics that were tackled through a combination of simulation and experimentation. The generation of excitation signals (mathematically, as opposed to physically) is covered first in subsection 5.1.2. Simple, non-parametric, stationary estimation of frequency response using Fourier transforms is then discussed in subsection 5.1.3. Finally, the use of adaptive filters for high-speed, dynamic response estimation, both for non-parametric and parametric spectral estimation, is introduced in subsection 5.1.4.

### 5.1.1 Linear system identification

The microwave spectral response of the resonator is the voltage wave response to voltage wave excitation as a function of frequency. This response is linear, *i.e.,* it is not a function of voltage itself, only frequency. Although certain materials

encountered in MRS measurement, such as ferrites, ferroelectrics, and superconductors, can show nonlinearity, the microwave excitation power can be normally reduced to a level where this is insignificant. Most resonant sensors, in fact, being high quality passive devices made from pure metallic conductors and the best dielectrics, are likely to be the among the most linear devices encountered in microwave measurement.

This linearity assumption is key to allowing multifrequency signals to be used as a measurement excitation. If a single frequency, $f_0$, is applied to a system with nonlinearity, the harmonics generated will be at multiples $2f_0, 3f_0$, *etc.* These can be rejected quite easily by an appropriate preselection filter at the input of the receiver. However, if a multifrequency signal is passed through a nonlinearity, distortion will be generated in-band (*i.e.,* at the same frequencies as used for excitation). This cannot be filtered away. Thus, the assumption of linearity of the microwave sensor system is not merely a mathematical abstraction but a necessary condition for the use of an excitation consisting of multiple frequency components.

Whilst the sensor itself may be linear, the system identification algorithm must take place in the digital domain. Therefore, there are several nonlinear components in the signal chain linking the digital world to the RF world and back again (see Figure 5.1). The following discussion of linear system identification therefore rests on the assumption that the distortion caused by these stages is mitigated or insignificant. The correction and avoidance of distortion using software techniques is discussed in later sections of this chapter.

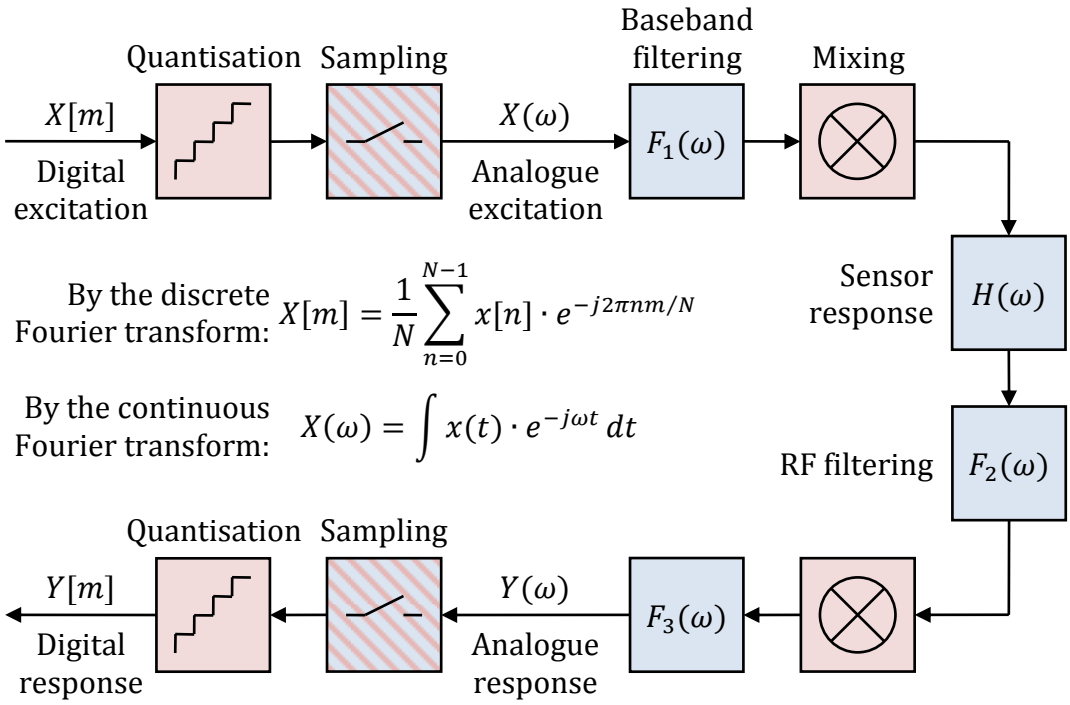



*Figure 5.1: Mathematical representation of the multitone system signal chain. Nonlinear components are shaded red, linear components blue. Sampling is not technically a nonlinear process, but causes spurious products to appear due to aliasing (akin to wrapping an infinite paper tape of continous domain spectra around a cylindrical drum of finite sampled space) that causes new spectral components to be generated in the same was as nonlinear distortion.*

From a mathematical perspective, an ideal mixing process does not alter the spectral characteristics of the excitation and response signals, just 'slides' them in frequency. Thus mixing stages can be replaced with a linear baseband model simply by redefining $H(\omega) \rightarrow H(\omega - \omega_c)$, where $\omega_c$ is the carrier, or LO frequency. Evidently, the assumption that both mixing stages are at exactly the same frequency of conversion must hold for this to be possible. Similarly, the assumption of ideal quadrature modulation and demodulation allows for perfect reconstruction of the upper and lower sidebands – leading to a complex signal representation of the entire bandpass excitation and response, with negative baseband frequency representing spectral content below the carrier frequency, and positive baseband frequency representing that above. This complex signal representation is only valid if the gain and phase offset of the two IQ mixers is

negligible. Thus, the analogue excitation and response signals, $x(t)$ and $y(t)$, are naturally represented as complex-domain signals with the real and imaginary components representing the *I* and *Q* channels of the transmitter and receiver.

Sampling and quantisation represent the last stages of the translation process to and from the digital world. Ideal sampling occurs for perfectly bandlimited signals, *i.e.,* when the baseband filtering $F_1(\omega)$ and $F_3(\omega)$ reject all signals with a frequency greater than Nyquist, $f_S/2$, where $f_S$ is the sampling frequency. Under these conditions, aliasing of out-of-band signals can be ignored and the sampling process, too, can be represented as a simple linear process: $x[n] = x(nT) = x(t)$ and $y[n] = y(nT) = y(t)$, where $T = 1/f_S$ is the sampling period. This highlights the importance of sampling synchronisation. For the excitation and response signals to be compared in the digital domain, the digital index, $[n]$, must be the same for both the transmitted and received signals. In the presence of a sampling time offset, for example, $x[n_2] = x(nT + \delta t) = x(t)$. The offset $\delta t$ is difficult to correct for unless it is static and an integer multiple of $T$. However, it can be tolerated if phase offsets are taken into account in later curve fitting procedures, although this may increase measurement error and system complexity unnecessarily. The simple and robust hardware solution of a shared sampling clock is therefore preferable.

Under this assumption, the sensor spectral response is represented in digital in a form known as the impulse invariant form – *i.e.,* its impulse response is similarly sampled, $s[n] = s(nT) = s(t) = \mathcal{F}^{-1}S(\omega)$. This preserves the shape of the spectral response and is the appropriate way of modelling the resonant sensor in the digital world. However, this also implies that the spectral response must be cascaded with that of the two ideal antialiasing filters and hence can only be known over a finite bandwidth $B = f_S$ – twice that of Nyquist band because of the unique positive and negative frequency information. In reality, these filters will have some pass-band ripple and finite transition bandwidth, which will alter the flatness of the response; however, since these effects are linear they are comparatively easily corrected.

The final quantisation process is the task of going from a finite number of integers to continuous voltages. This is inherently nonlinear and will result in distortion. However, an 'ideal' quantiser model approximates the resulting very high-order, low-level distortion as an additive white noise source (quantisation noise) that is at a constant level of about one or two units of the quantisation interval, known as least-significant bits (LSBs). Along with any other thermal noise in the baseband circuitry, this can thus be represented as two additive random noise sources, $\nu_1$ and $\nu_2$.

The resulting idealised, linear mathematical model of the multitone hardware system is now simple enough to apply linear system identification theory and gain an insight into the use of multifrequency excitation and response signals in estimating the spectral response of an RF system from the digital world. Later sections of this chapter will deal with the various non-ideal behaviours of the single chain that cause the real-world multitone system to deviate from this mathematical ideal.

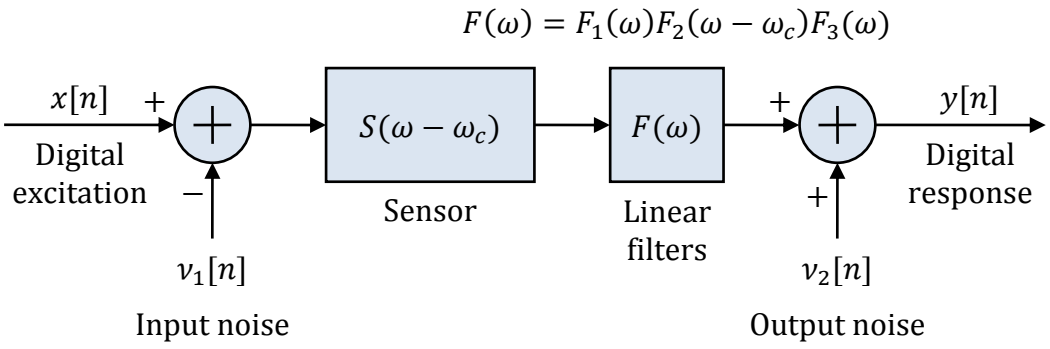



*Figure 5.2: Simplified stochastic model of the multitone system.*

### 5.1.2 FFT-based sensor spectral response estimation

In this model and discussion, the spectral representation of the sensor and system components is used extensively. However, as mentioned in the Theory chapter, the sensor is a time-varying system that cannot be represented by a Fourier transform exactly, as Fourier transforms only exist for strictly stationary signals. Fortunately, in the digital domain, use can be made of the inherently time-limited nature of the

discrete Fourier transform (DFT). This is because the DFT can be broken down into an equivalent series of operations that make no assumptions of stationarity. In effect, the discrete Fourier transform performs a digital downconversion at each frequency point, consisting of quadrature digital mixing (by a complex LO), low-pass filtering (by a moving average filter) and finally decimation (rate reduction) by a factor *N,* the number of samples used for the DFT estimate. In effect, the DFT mixes a composite signal down to DC from a particular discrete frequency 'bin' at $f_m = mf_S/2N$, where $-1 \leq mN < 1$. Thus, the quadrature DC value (or mean) is extracted by the summation over *N* samples. If every other frequency component of the signal is periodic in the window of *N* samples then they will have a mean of zero and thus cancel out completely. This just leaves an estimate of the magnitude and phase at that frequency, wholly analogous to the homodyne quadrature readout system. This is another way to look at the spectral leakage problem, and highlights the importance of using excitation signals that are periodic in the digital analysis window. Signals that are not periodic will not cancel out in the averaging process and instead will 'leak' across the estimates of all spectral components.

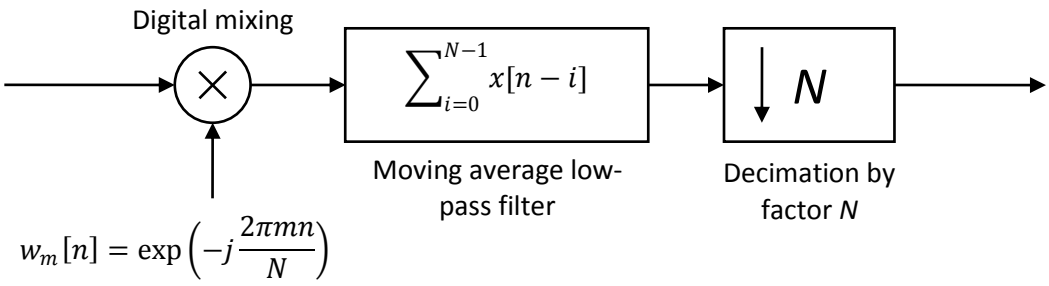



*Figure 5.3: Representation of the DFT as an on-going time-domian digital downconversion process consisting of digital mixing, low-pass filtering and decimation.*

To avoid aliasing in a decimation process, it is assumed that the input signals to the decimator are perfectly bandlimited to contain no frequency content above $f_c = f_S/N$. If this is not the case, aliasing may occur. This situation can be greatly improved by replacing the moving average (MA) filter with an alternative finite impulse response (FIR) low pass filter (Figure 5.4). This filter is four times longer, but has a closely matched response to the MA whilst having much higher alias

rejection. Use of such a filter avoids the guesswork of the FFT approach in dynamic signal environments, preventing unexpected inaccuracies arising from unexpectedly high transients, *etc.* When only a few frequencies are required, the extra computational burden of this technique is not an issue.

The DFT estimate is therefore still valid in the case of a time-varying system, such as the resonant sensor, as long the desired response rate of the sensor change is less than $f_S/N$. In the multitone system, typical parameters are $f_S = 100 \cdot 10^6$ and $N = 64$, giving a cut-off of $f_c \approx 1.6\,MHz$. In all of the practical experiments carried out, the rate of change of the resonant sensor spectral response is several orders of magnitude less than this. Therefore, use of the conventional DFT is still valid. However, this also highlights a limitation of the DFT method. Frequency resolution, *i.e.,* the number of frequency samples, $N$, and the cut-off for dynamic changes, $f_c$, are inextricably linked. As discussed in the last subsection, adaptive filters provide a neat way of allowing a dynamic trade-off to be reached instead. This means that the maximum measurable rate of change of sensor characteristics can be decoupled from the achievable frequency resolution.

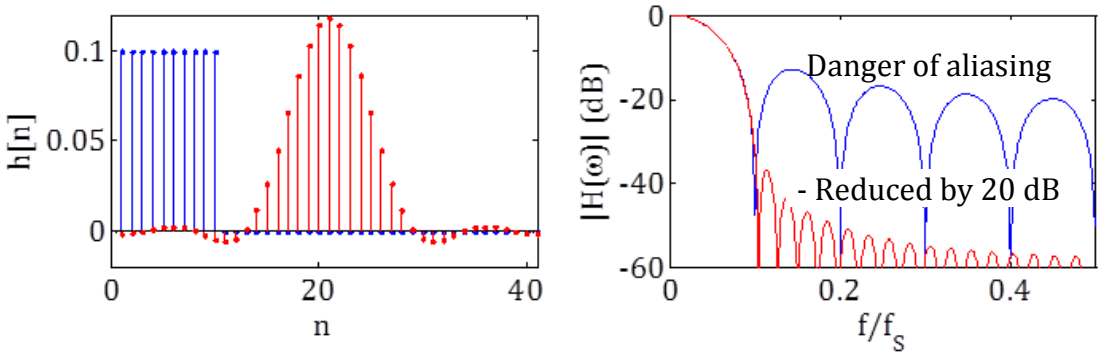



*Figure 5.4: $N = 10$ MA LPF impulse response (left) and magnitude response (right) in blue, compared to a least-squares FIR filter design in red, matched to the MA response in the pass band but offering much higher alias rejection; i.e., suitable for use with dynamic signals.*

The simplified model of the multitone system also contains random, or stochastic, signals on both the input and output. Stochastic signals do not have a strictly defined Fourier representation, as they not square-integrable: $\int_{-\infty}^{\infty}|x(t)|^2 dt \not< \infty$.

Therefore, their spectra must be derived through the Wiener-Khinchin theorem [1] that relates the spectrum to the statistical expectation of the autocorrelation function

$$S_{xx}[m] = \frac{1}{N}\sum_{n=0}^{N-1} r_{xx}[n] \cdot e^{-j2\pi nm/N}, \quad r_{xx}[n] = \mathrm{E}(x[n] \cdot x^*[n-k]), \tag{5.1}$$

where $S_{xx}[m]$ is the power spectrum and $r_{xx}[n]$ is the expected autocorrelation at time sample $n$. $S_{xx}[m]$ is always real. The cross-power spectral density between two signals, $S_{xy}[m]$, (complex, in general) can be similarly derived:

$$S_{xy}[m] = \frac{1}{N}\sum_{n=0}^{N-1} r_{xy}[n] \cdot e^{-j2\pi nm/N}, \quad r_{xy}[n] = \mathrm{E}(x[n] \cdot y^*[n-k]). \tag{5.2}$$

Using these identities, it is possible to derive two different ways of estimating the spectral response (*i.e.,* transfer function) of the combined system and sensor, $H(\omega) = S(\omega - \omega_c)F(\omega)$, at the discrete frequencies, $\omega - \omega_c = m/NT$, using digital signal processing: [2]

$$H_1[m] = \frac{S_{xy}[m]}{S_{xx}[m]}, \qquad H_2[m] = \frac{S_{yy}[m]}{S_{yx}[m]}. \tag{5.3}$$

As it stands, it is not possible to evaluate these expressions directly as the definition of the statistical expectation of the correlation functions contains an average over all time. However, using the linearity property of the DFT, the auto- and cross-correlation operations can be taken outside of the DFT summation, and an approximation to the transfer function of the sensor can be derived using the statistical average over $M$ individual DFT estimates of the excitation, $X[m]$, and response, $Y[m]$. The estimates must be the same length, thus the number of samples used for the excitation is the same as the number of samples acquired for each response. These estimates can be efficiently generated using the fast Fourier transform (FFT) algorithm:

$$\widehat{H}_1[m] = \frac{\frac{1}{M}\sum_{k=0}^{M-1} Y[m,k]\cdot X^*[m,k]}{\frac{1}{M}\sum_{k=0}^{M-1} X[m,k]\cdot X^*[m,k]}, \qquad \widehat{H}_2[m] = \frac{\frac{1}{M}\sum_{k=0}^{M-1} Y[m,k]\cdot Y^*[m,k]}{\frac{1}{M}\sum_{k=0}^{M-1} X[m,k]\cdot Y^*[m,k]}. \tag{5.4}$$

Comparing both these estimates to the ideal noise-free case, it is possible to see that the presence of noise at both the input and output leads to bias: [2]

$$\widehat{H}_1 \approx H\cdot\left(1+\frac{S_{v_1v_1}}{S_{xx}}\right)^{-1}, \qquad \widehat{H}_2 \approx H\cdot\left(1+\frac{S_{v_2v_2}}{S_{yy}}\right), \tag{5.5}$$

where $H$ is the 'true' sensor transfer function, $S_{v_1v_1}$is the noise spectral density at the input and $S_{v_2v_2}$is the noise spectral density at the output. The fractions $S_{v_1v_1}/S_{xx}$ and $S_{v_2v_2}/S_{yy}$ can be seen to be the input SNR and output SNR, respectively. Reducing the noise on either the input or output (by using a higher resolution digital/analogue converter and better quality circuits), and choosing the appropriate version of the estimates, $\widehat{H}_1$ for low transmitter noise and $\widehat{H}_2$ for low receiver noise, will reduce the bias of these estimates. Since $S_{xx}$ is generally spectrally 'flat' and time-invariant when using a static excitation signal, it is clear that $\widehat{H}_1$ is most appropriate estimate, as the resulting bias is constant across frequency and thus will have minimal effect on the derived sensor eigenvalue estimates.

To reduce bias, as demonstrated by the simple computer experiment in the Hardware chapter, vector averaging is a preferred to averaging in the power domain, and this applies equally to spectral estimation. Thus, a simple, efficient, and low bias method of estimating the sensor transfer function can be derived:

$$\widehat{H}[m] = \frac{1}{M}\sum_{k=0}^{M-1}\frac{Y[m,k]}{X[m,k]}. \tag{5.6}$$

This is equivalent to the vector average of $M$ individual estimates of the complex transfer function. A powerful feature of this method of estimating the sensor transfer function is that only the specific frequency bins that contain excitation

signal tones need be evaluated. Thus, out of the full vector of the $kth$ response signal $\mathbf{Y}[k]$ one can select an arbitrary subset $\mathbf{Y_2}[k]\ \forall\ m \subseteq \{0..N-1\}$, an operation equivalent to an ideal 'brick-wall' filter. This is an error-free process as long as all signals are periodic in the original time window and spectral leakage is not present.

Furthermore, this algorithm may be simplified with the knowledge that the excitation signal spectrum does not change with time, hence may be taken outside of the average. This avoids having to do an explicit division operation at fast rate:

$$\widehat{H}[m] = \frac{1}{MX[m,k]} \sum_{k=0}^{M-1} Y[m,k] \quad \forall\ m \subseteq \{0..N-1\} \tag{5.7}$$

To implement this approach in an online system, the approximately stationary assumption can be used to perform the above algorithm continuously over repeated blocks of $N$ samples. When continuously supplied with data, this algorithm outputs a new spectral estimate at a rate of $f_S/MN$, where $N$ is the length of the FFT used (typically quite short *e.g.,* 64 samples), and $M$ is the number of averages performed (a typical value used is 128). At an input sample rate of 100 MS/s, these settings would give a spectral response estimate at a rate of about 12 KHz.

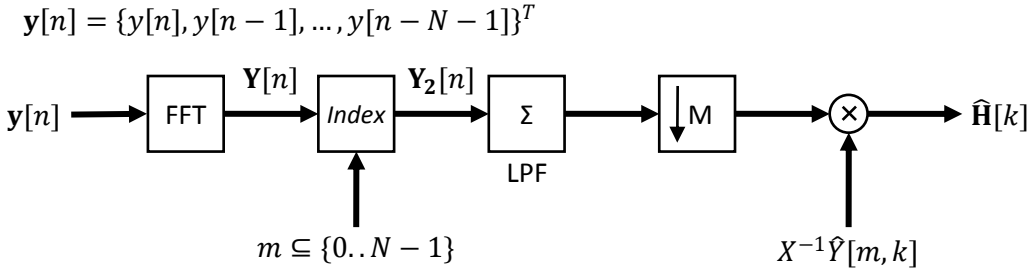



*Figure 5.5: Diagram of the block FFT algorithm for estimating the spectral response of the sensor as an online process.*

The variance of this spectral estimate is given by

$$\sigma_{\widehat{H}}^2[m] = E\{(\widehat{\mathbf{H}}[m] - E\{\widehat{\mathbf{H}}[m]\})(\widehat{\mathbf{H}}[m] - E\{\widehat{\mathbf{H}}[m]\})^*\} \quad 5.8$$

Using a first order linear approximation, this gives an estimate for the variance of the spectral estimate of:

$$\sigma_{\widehat{H}}^2[m] \approx \frac{2}{M}\left(\left|\widehat{H}[m]\right|^2 \frac{|\mathrm{N}_1[m]|^2}{|\,\mathrm{X}[m]\,|^2} + \frac{|\mathrm{N}_2[m]|^2}{|\,\mathrm{X}[m]\,|^2}\right), \quad 5.9$$

where $|\mathrm{N}_1[m]|^2/|\mathrm{X}[m]|^2$ is the ratio of the input noise power to excitation signal power at frequency $m$, and $|\mathrm{N}_2[m]|^2/|\mathrm{X}[m]|^2$ is the ratio of output noise power to excitation signal power. This analysis shows that, as expected, averaging improves the SNR of the estimate, $\langle\widehat{H}[m]\rangle/\sigma_{\widehat{H}}[m]$, by a factor of $\sqrt{M}$.

Equation 5.9 predicts that the effect of output noise is independent of the sensor itself – and therefore of any sample perturbation. However, the effect of input noise is proportional to the sensor's power transmission. This will cause the derived eigenvalue estimate to have bandwidth-dependent noise, because the change in bandwidth is inversely proportional to insertion loss. In other words, the sensor system will have heteroscedasticity. This can be a problem, as common curve fitting and statistical analysis techniques often make the assumption of constant variance, and may therefore be invalidated.

Heteroscedasticity can be reduced by ensuring that output noise is always dominant. In the hardware implementation, this was indeed the case, as a 14-bit converter at the receiver and a 16-bit converter at the transmitter. To investigate the performance of the FFT spectral response estimation algorithm for various input/output noise conditions, numerical Monte-Carlo simulations were performed using additive white Gaussian pseudorandom noise and analysed over 100,000 trials.

The simulations are in agreement with Equation 5.9, with the variance due to input noise being Lorentzian, *i.e.* $\propto |\mathbf{H}|^2$ as predicted, in both $\Re\widehat{\mathbf{H}}$ and $\Im\widehat{\mathbf{H}}$. For the power transmission estimate, the variance is between Lorentzian and squared-Lorentzian

depending on the relative input and output noise levels. Output noise also causes a significant decrease in the precision of the phase estimate away from resonance.

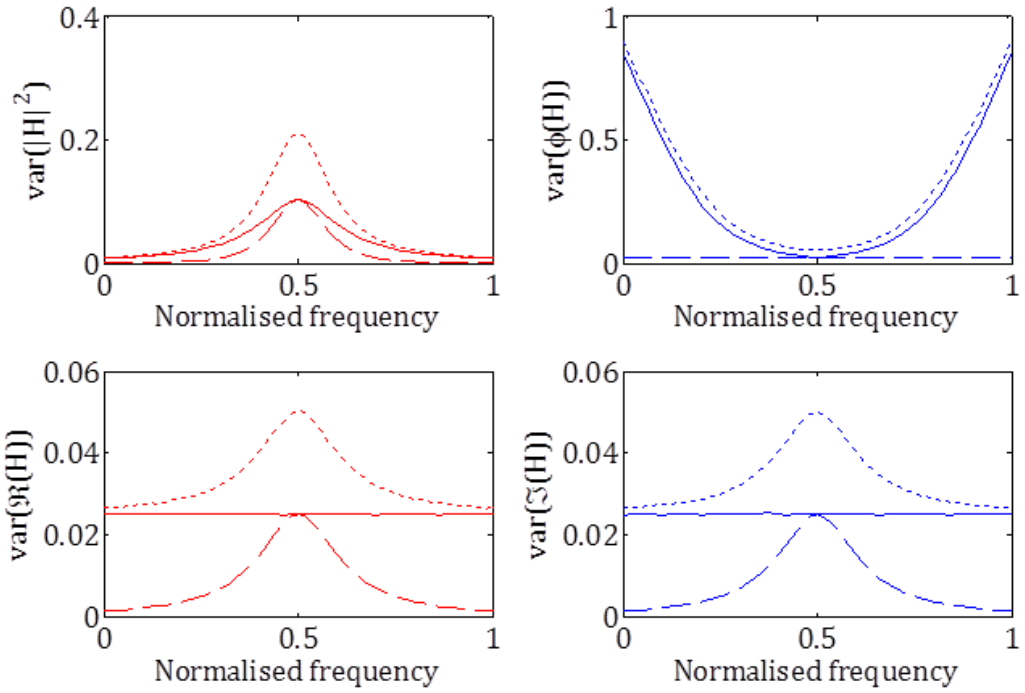



*Figure 5.6: Variance of* ***H*** *estimates. Continuous lines are for output SNR of 6 dB with no input noise, dashed for input SNR of 6 dB with no output noise, and dotted for an SNR of 6dB at both the input and output.*

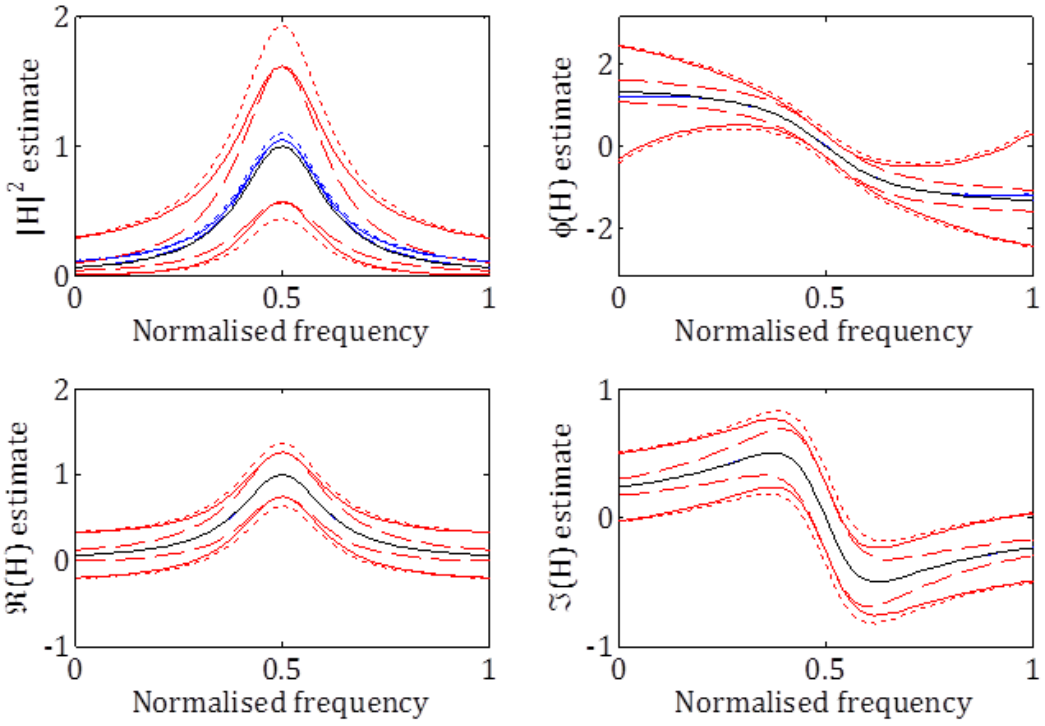



*Figure 5.7: Estimates of **H** in various complex mappings. The black line shows the true value. Blue lines are the estimate ensemble mean, red are the 5% and 95% ensemble percentiles. Lines are as Figure 5.6.*

The estimated relative bias of the sensor response estimate is defined as $\langle(\widehat{\mathbf{H}} - \mathbf{H})/\mathbf{H}\rangle$. This figure was also computed for the three different combinations of input and output noise. The results are shown in Figure 5.8. The results confirm that, with a finite number of FFT averages (10 in this case) the power transmission response still has residual bias. Furthermore, this bias is frequency dependent when noise is present at the output of the resonant sensor (the very situation preferred on homoscedasticity grounds) appearing to be roughly parabolic with frequency around resonance.

For power-domain eigenvalue estimation methods, such as Lorentzian curve fitting, this means that the bandwidth will be overestimated. However, the added noise level at the output should be constant, regardless of sensor perturbation, thus differential bandwidth measurement should cancel out this bias and not affect the accuracy of any derived measurements. Even if bias on absolute bandwidth estimates is a problem, at realistic system noise levels and with a reasonable

number of FFT averages (the bias decreases linearly with number of averages), the level of this bias should be insignificant. Evidently, however, the issue can be avoided altogether by using $\Re\{\widehat{\mathbf{H}}\}$ and $\Im\{\widehat{\mathbf{H}}\}$ to extract the eigenvalue estimate.

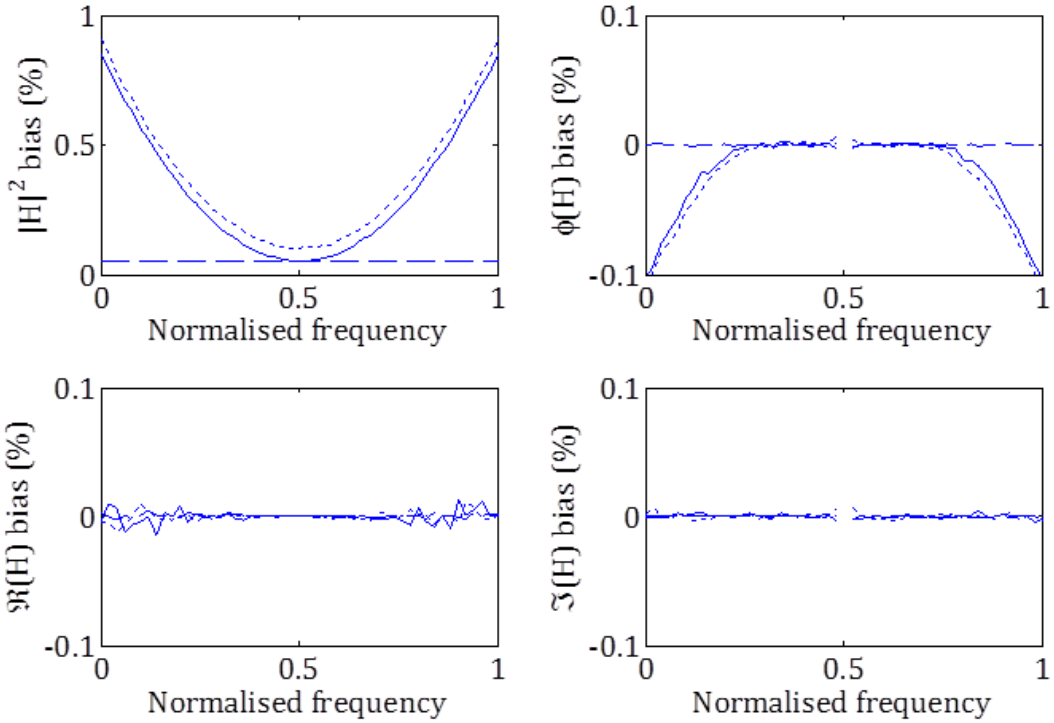



*Figure 5.8: Bias of the simulated sensor transfer function estimates for various complex mappings. Continuous lines are for an output SNR of 6 dB with no input noise, dashed for an input SNR of 6 dB with no output noise, and dotted for an SNR of 6dB at both the input and output of the resonant sensor. 10 averages were used in the algorithm.*

To see how the variance in sensor transfer function estimates impacts on eigenvalue estimates, the mathematical inverse relationship, Equation 3.81, can be used. This is the most efficient and lowest bias eigenvalue estimator, provided amplitude estimates are accurate, and the resonator constants $K_1$ and $K_0$ are known exactly – unlikely in practice, but this direct inverse is still useful in providing a benchmark.

The results of the Monte Carlo simulations (Figure 5.9) show that with only input noise, the best-case SNR (*i.e.,* at the optimum frequency) of the eigenvalue estimates is roughly equal, at about 85 dB on average – a SNR gain of 25 dB over

the spectral estimate used to derive them. However, for output-only noise, the SNR gain is much higher, up to 100 dB for small changes, but varies significantly with the perturbation (*i.e.,* having significant heteroscedasticity). This is the opposite of what Equation 5.9 seems to predict for the transfer function estimate itself. These results can be used to derive two recommendations. Firstly, better sensor precision is reached by concentrating on minimising the noise at the transmitter in preference to at the receiver, as a proper eigenvalue estimator should have better output noise rejection. Secondly, if the output noise at the receiver causes undesirable heteroscedasticity, then this can be alleviated by making the noise at the transmitter more dominant overall.

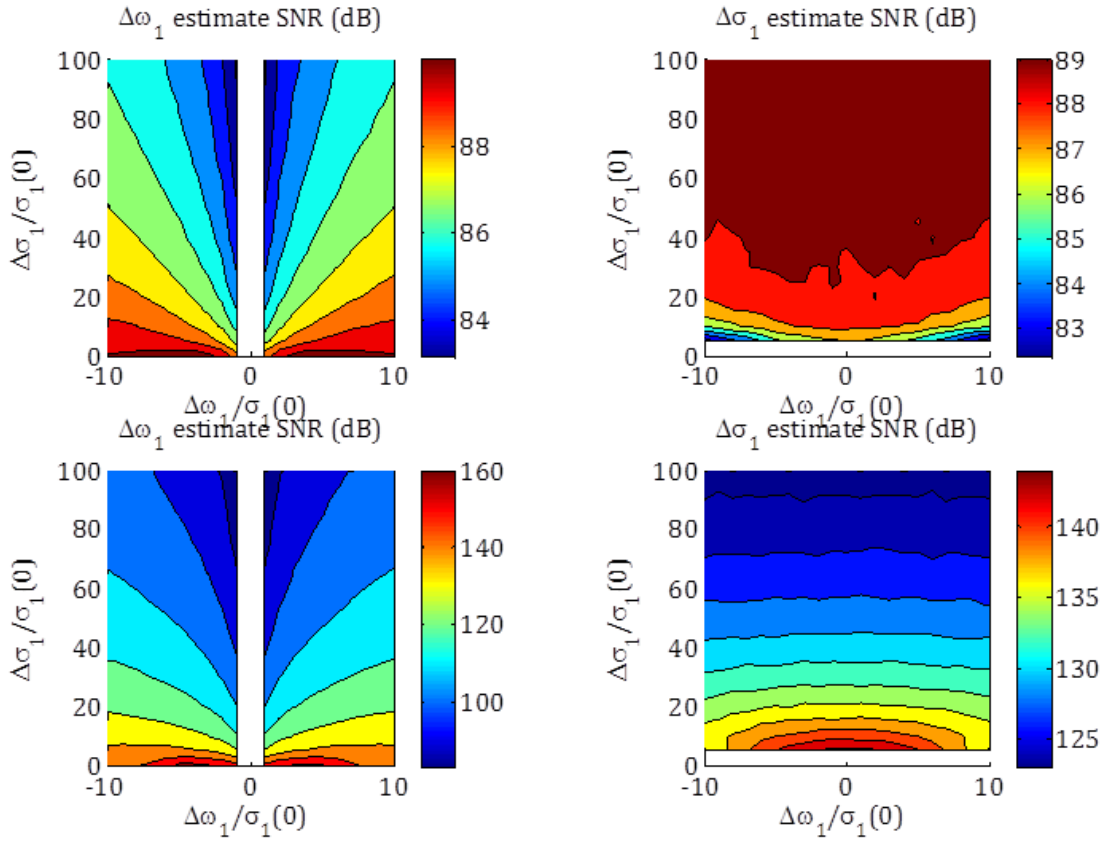



*Figure 5.9: Maximum SNR of the sensor eigenvalue estimate as a 2D function of bandwidth change $\Delta\sigma_1$ and resonant frequency change $\Delta\omega_1$. The top row are for an input SNR of 60 dB with no output noise, and the bottom row are for an output SNR of 60 dB with no input noise. The simulated resonator had $Q_1(0) = 5000$ and $f_1(0) = 1$ GHz, no additional averaging was used.*

These results suggest that very high resolution ADC and DACs are not necessary for high precision MRS measurements, as through a combination of spectral averaging and the inherent SNR gain of the eigenvalue estimation process much of the quantisation noise should be rejected. Indeed, the high SNR gain of the estimation process is consistent with that found in the long-term stability measurements (see Section 4.6), as both the VNA and the multitone system are receiver-noise dominated, having a receiver noise floor of around 90 dB yet showing a peak SNR of around 180 dB in resonant frequency estimation.

However, this does not mean cheap 8-bit ADCs and DACs could necessarily replace for the high-resolution devices used here. Nonlinear distortion in the ADC and DAC

response limit the improvement possible with this method as, in actuality, quantisation noise is not truly random but consists of a large number of deterministic spectral components known as spurious signals. Spurii will not be reduced by averaging and will cause a source of bias on the spectral response estimate. One simple measure that could reduce this effect is to inject a low-level random 'noise' signal, known as dither. This can applied digitally in the case of the DAC, and in the case of the ADC, if the normal noise figure of the receiver baseband circuitry is not sufficient by itself, it can be supplemented using an analogue circuit such as a 'noise diode' amplifier. Dither provides a randomly changing input to the quantiser, causing the spurii to be also randomly time varying and thus more amenable to reduction by averaging. Clearly, the accuracy advantage gained by this method must be balanced by the precision degradation. However, in most cases, MRSs have plenty of precision and accuracy is really the limiting factor, making this an attractive option.

Measurements of the digitizer used in the second multitone system implementation indicated that it has a sufficiently high level of random noise present at its inputs already due to the input amplification stages, meaning additional dither was not a necessity. However, other sources of nonlinearity serve to limit the maximum accuracy that can be obtained in sensor spectral estimates well before the effects of random noise become significant. This is discussed further in Section 5.2.

### 5.1.3 Multifrequency excitation design

The above subsection showed how the sensor transfer function can be estimated given an arbitrary finite-length digital excitation, **X**. Since the excitation is arbitrary (meaning that there should be no need to alter it), and of finite length equal to the FFT used in the estimation algorithm, it can just be repeated continuously by the DAC using a simple circular memory buffer. This makes its generation simple. So what should the excitation be? It should definitely contain multiple frequencies, allowing the spectral estimate to be made across the resonator bandwidth. This then replaces the need to sweep the signal or otherwise alter it, and, if the signal

covers sufficient bandwidth, avoids any necessity to change the local oscillator frequency. This means a simple fixed-frequency oscillator can be used rather than an expensive synthesiser. The choice thus comes down to how best to assign those *N* digital numbers of the excitation signal in some optimum way.

The performance of a given spectral estimation method can be measured in terms of the spectral estimate variance, $\sigma_{\hat{H}}^2$, that it gives under stationary signal conditions. Furthermore, to introduce a measure of the efficiency of a given method, the spectral variance can be normalised to a given time, or equivalently a given number of samples. Thus, the precision-per-second is a good measure by which to compare various excitation signals that can be used for spectral response estimation. Because the FFT used at the spectral estimate side must be the same length as the excitation, it therefore makes sense to normalise to this length. The performance measure is therefore the minimum variance $\sigma_{\hat{H}}^2$ that can be achieved in one N-sample period of the excitation signal.

A figure of merit used to compare different signals is the crest factor [2]. The crest factor normalised to that of a sinewave (*i.e.,* $\sqrt{2}$ ), $C_n$, for a discrete signal $x[n]$ is given by:

$$C_n = \frac{\max\ |x[n]|}{x_{rms}\sqrt{2}}, \quad x_{rms} = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}|x[n]|^2} \qquad 5.10$$

This equation is often converted to dB, in which case the lowest crest factor physically obtainable is $-3.01\ dB$ relative to a sinewave. This is the case for a DC signal or square wave excitation. Due to the requirement that the excitation fit in the maximum full-scale range (FSR) of the generating DAC, it is normally the case that the maximum amplitude, $\max|x[n]|$ is fixed. Therefore, the average power of the excitation signal, and therefore its SNR, is related fundamentally to the crest factor:

$$SNR_x = \frac{x_{rms}^2}{v_{rms}^2} = \frac{\max\ |x[n]|}{2\ C_n^2\ v_{rms}^2} \propto \frac{1}{C_n^2}, \tag{5.11}$$

as both $\max|x[n]|$ and the noise power $v_{rms}^2$ are constants for a given hardware configuration. Substituting Equation 5.12 into Equation 5.9, the $\widehat{\mathbf{H}}$ variance is:

$$\sigma_{\hat{H}}^2[m] = \frac{4C_n^2}{NM}\left(\frac{|H[m]|^2\ |\mathrm{N}_1[m]|^2 + |\mathrm{N}_2[m]|^2}{\max|x[n]|}\right). \tag{5.12}$$

as, from Parseval's theory [3],

$$x_{rms}^2 = \frac{1}{N}\sum_{n=0}^{N-1}|x[n]|^2 = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1}|X[m]|^2 = N|X|^2, \tag{5.13}$$

assuming the excitation is spectrally flat. The SNR of the transfer function estimate is, therefore

$$SNR_{\hat{H}[m]} = \frac{\sqrt{NM\ \max|x[n]|}}{2C_n}\left(\frac{H[m]}{\sqrt{|H[m]|^2\ |\mathrm{N}_1[m]|^2 + |\mathrm{N}_2[m]|^2}}\right). \tag{5.14}$$

The total number of measurement samples in this estimate is $NM$, and the comparative excitation signal figure of merit is the precision-per-sample, *i.e.,* $\sigma_{\hat{H}}^2[m]/NM$, thus it is independent of these numbers. Furthermore, we have no control over the term in brackets as it depends only on the hardware and the sensor. Therefore, the only parameter that can reduce the spectral estimate variance independently of acquisition time is the excitation crest factor, $C_n$. Since increasing the measurement time is undesirable, it is imperative that the crest factor is minimised for whatever excitation signal is chosen. This precludes multifrequency excitation signals such as an impulse train, which have very high crest factors (despite being relatively easy to generate). Given that the comparison of performance for excitation signals basically comes down to this one figure of merit, the choice of excitation signal must largely be dictated by practical considerations – just as long as it has the lowest crest factor possible.

Many signals that do have reasonably low crest factors, such as stepped sine waves or periodic 'chirp' signals of periodically varying frequency, have undesirable additional frequency components extending over a wide bandwidth, these can interfere when aliased or mixed with the desired excitation by intermodulation distortion. Many excitation signals do not allow arbitrary control of amplitude, or are not periodic in the FFT window length, leading to spectral leakage (*e.g.,* pseudorandom noise). One of the best candidate excitation signals are maximum length binary sequences (MLBS). Being binary, these only need a one-bit DAC to be generated. However, these signals also have disadvantages: whilst excellent as a baseband excitation signal, they lose many of their desirable properties once modulated. They also permit no control over the location or relative amplitude of its spectral components.

Therefore, after much consideration into the practicalities of possible excitation signals, a multisine excitation signal was chosen. Multisine signals and perturbation signals in general are covered in the excellent text by Godfrey [2]. This excitation signal has a number of practical and theoretical advantages, as summarised in Table 5.1.

*Table 5.1: Advantages of a multitone excitation signal*

| Advantage | Description |
|---|---|
| Multifrequency | Multi-point spectral characterisation is essential when both bandwidth and resonant frequency are required or the change in resonant frequency is large and/or rapid |
| Persistently exciting | Each frequency component is static as a function of time, therefore avoiding resonant transient response. This is not the case with stepped sine (*i.e.,* VNA), impulse, *etc.* |
| Deterministic | The excitation signal is known without having to explicitly measure it or rely on statistical methods (*e.g.,* noise excitation) |
| Arbitrary/flat amplitude | Excitation spectrum can be of arbitrary amplitude. This is useful in compensating filter roll-off, as it ensures uniform SNR. MLBS cannot have arbitrary amplitude spectra |
| Low crest factor | Optimised multisine signals have a crest factor only slightly greater than a single sine wave, much lower than, *e.g.,* pseudorandom noise |
| Broadband | In theory, a multisine can cover a bandwidth as wide as the Nyquist limit of the generating DAC. Analogue modulation, *e.g.,* FM is limited by LO PLL bandwidth |
| No wasted spectrum | With ideal generation hardware, a multisine has no spectral components other than those directly used for measurement |
| Zero spectral leakage | As long as each sinusoidal component is periodic, *e.g.,* in 256 samples, spectral leakage is zero. Therefore, window-free FFT processing can be used without error |
| Flexible number of samples | No limit on the number of samples, *i.e.,* period can be a power of two as FFT processing is most efficient. This is not the case with MLBS, which have period $2^N - 1$ |
| Trivial digital modulation | Digital modulation is achieved simply by selecting tones from around some offset frequency. RF modulation also works as it is band-limited and robust to distortion |

The equation of a discrete complex-domain multisine signal is

$$x[n] = \sum_{i=0}^{P-1} A_i \cdot e^{j(2\pi f_i nT + \phi_i)}, \qquad n = 0..N-1, \tag{5.15}$$

where $P \leq N$ is the number of tones of the multisine signal. In general, a multisine can therefore consist of arbitrarily spaced tones of arbitrary amplitudes; however,

equally spaced, flat amplitude signals are most useful for excitation signals. In the digital world, when the number of cycles of each sinusoid is an integer, $N/f_i T \in \mathbb{Z}$, then spectral leakage is zero and the entire multisine signal consists entirely of individual discrete line spectra. This is known as a multitone spectrum. An easy way to generate a multitone signal and ensure this condition is always met is to use the inverse DFT:

$$x[n] = \mathcal{F}^{-1}\{\mathbf{A}e^{j\mathbf{\Phi}}\} = \sum_{m=0}^{N-1} A[m]e^{j\phi[m]} \cdot e^{j2\pi nm/N}, \qquad m = 0..N-1 \qquad 5.16$$

which can be seen to reduce to Equation 5.15 when $f_i = mT/N$. It is trivial to specify a particular tone subset, $i \in \{0..P-1\} \subseteq \{0..N-1\} \ni m$, by setting a particular $A[m]$ to zero. Evidently, in the limit of a long inverse FFT (IFFT) arbitrary frequency resolution can be achieved. However, it is more efficient to use few tones and space them uniformly and as closely as possible, keeping the FFT short and computationally efficient. Matching the length of the generating IFFT to the FFT used on the acquired waveforms will ensure that zero spectral leakage is preserved without wasting spectral resolution with an unnecessarily large number of samples.

### 5.1.4 Digitally modulated multitone signal

Digital modulation is trivial for a multitone. Tones are simply centred on a particular carrier frequency. There is a good reason for doing this. Digital modulation at one-quarter sampling frequency is often used to implement direct downconversion receivers and likewise transmitters because of its simplicity; the digital IQ mixing by a wave at $f_S/4$ is then trivially implemented by multiplication by the cosine and sine sequences $[1, 0, -1, 0]$ and $j[0, 1, 0, -1]$ [4]. Apart from its simplicity, this choice of modulation frequency also minimises intermodulation distortion (IMD).

IMD is caused by nonlinearities in both the baseband and RF parts of the system. Given a memoryless nonlinearity, this is represented by the power series

expansion, $y[n] = \sum a_i x^i[n]$. This is a particularly sensitive issue for multitone signals as IMD causes inter-mixing of adjacent tones, making their amplitudes interrelated in a nonlinear way. For $N$ frequency components, $f_{i\,=\,1..N}$, intermodulation products are created at frequencies defined by $f_{IMD} = k_1 f_1 + k_2 f_2 + \cdots + k_N f_N$ where $k_i$ can be any integer, both positive and negative. The IMD product's order is defined $O \triangleq \sum_{i=1}^{N} |k_i|$. For example, the classic two-tone third-order products that lie in-band, often the dominant source of nonlinearity caused by the power amplifier in an RF transmitter, are given by $f_{IMD3,1} = 2 \cdot f_1 - 1 \cdot f_2$ and $f_{IMD3,2} = 2 \cdot f_2 - 1 \cdot f_1$. For a multitone excitation where all components are periodic in the FFT window, every one of these IMD products will also fall directly on, or be aliased to, an in-band FFT bin. Thus, nonlinearity cannot cause spectral leakage in itself.

Generally, only odd-order RF nonlinearities cause in-band distortion as even order distortion products are created at much higher RF frequencies and are filtered away. However, baseband nonlinearities can cause IMD of any order – *e.g.,* those caused by the IQ mixer at the transmitter have significant second and third order baseband nonlinearity. Digital modulation can avoid certain IMD products, as shown in Figure 5.10. Although the number of intermodulation products does not represent their absolute amplitudes, this graph does indicate that modulation around $f_s/4$ causes many of distortion products to fall outside of the signal band and thus have minimal impact after filtering (*i.e.,* tone indexing).

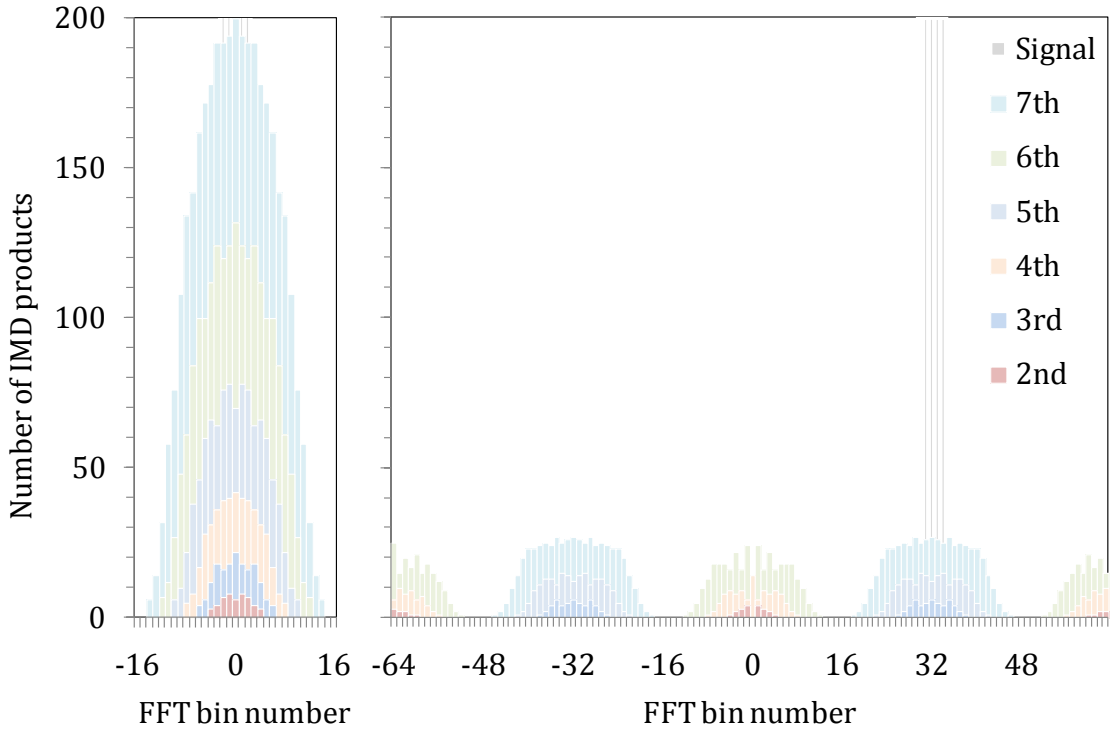



*Figure 5.10: Comparison of the number and location of baseband-generated intermodulation (IMD) products of a 5-tone multisine signal located around DC (left) to the same signal located around $f_s/4$ (right).*

Digital distortion can also be reduced with other tone location methods, for example by choosing all tones from a sequence of prime multiples [5]. This minimises the probability that IMD products will fall on tone locations. However, this vastly increases the length number of samples required for the excitation, as the number of cycles required becomes quite large in order to ensure roughly equal tone spacing. This, in turn, would decrease the maximum readout rate if an FFT method were employed. If nonlinear distortion does become a limiting factor, turning to these signals would improve performance at the expense of measurement rate, but would not require any alteration to the physical hardware.

### 5.1.5 Multitone signal crest factor optimisation

As indicated above (Equation 5.12), the variance of the transfer function estimate is critically dependent on the crest factor. For a particular excitation signal, therefore, this parameter must be minimised. For a multitone signal, the crest

factor is highly dependent on the relative phase of each component. If all phases of a large number of tones are all set to the same value, the resulting time-domain signal tends towards a series of periodic impulses having an extremely high crest factor.

An analytical approach to ensuring a low crest factor was found by Schroeder [6] for equally spaced multitone signals. This defines the phases of the multisine components according to the simple progression:

$$\phi[m] = \frac{\pi}{P}[m(1-m)], \quad m = 0 \ldots P-1 \tag{5.17}$$

However, this solution is non-optimal for digitally modulated or non-uniformly spaced multitone signals. One option first proposed by Van der Ouderaa *et al.* [7] is to use an iterative time-domain clipping approach, where the multisine is repeatedly passed through a time-domain level clipping function before transforming back into the frequency domain. Only the resulting clipped tone phases are kept, and the process is repeated. This method was adapted for the multitone signals generated here. Due to the sensitivity of the convergence of this algorithm to the exact level use for the clipping, the algorithm was slightly modified, introducing a 'soft clip' hyperbolic tangent function and a stochastic element to the clip level.

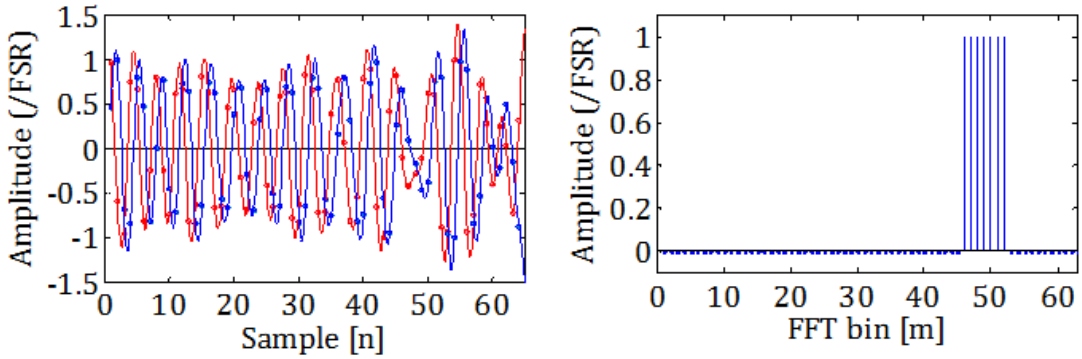



*Figure 5.11: Example of a crest factor optimised, 7-tone digitally modulated multitone in time (left), and in the frequency domain (right). Red are real samples, blue imaginary. The actual digital samples are shown as points, their continuous-time representation as a curve.*

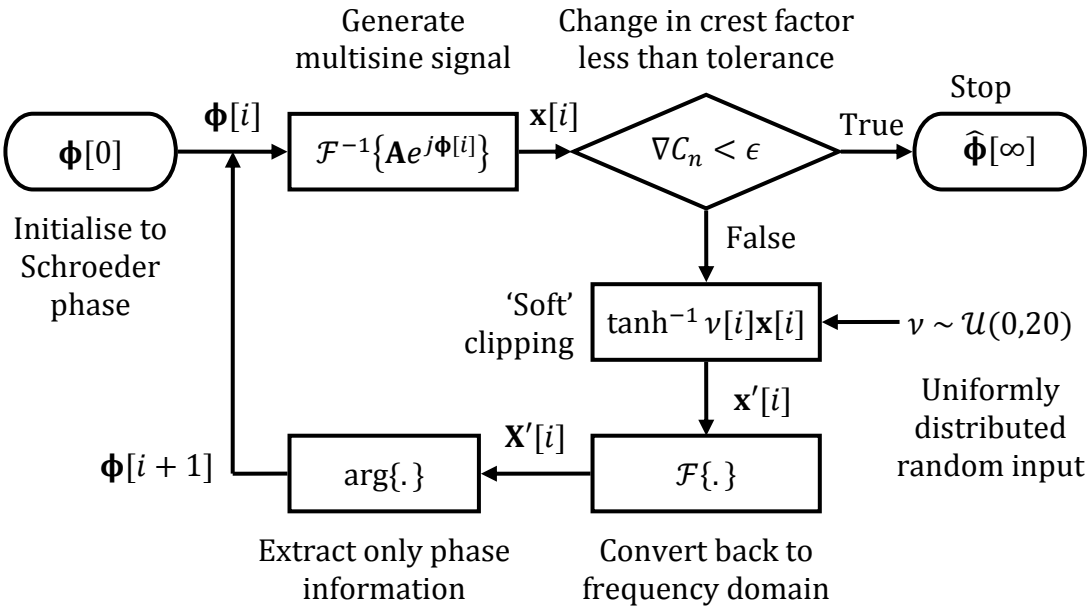



*Figure 5.12: Modified stochastic crest factor optimisation algorithm*

To explore further the minimisation of multitone crest factor, an optimisation algorithm was used. The phase vs. crest factor space is high-dimensional for larger numbers of tones, and the performance function contains an extremely large number of local minima. This makes it a difficult challenge for simple optimisation algorithms such as Gauss-Newton *etc*. A relatively recent and powerful stochastic optimisation algorithm known as particle swarm optimisation (PSO), inspired by the 'collective intelligence' of swarming bees, [8] is ideally suited to this type of

problem as it is well-suited to problems with large numbers of local minima due to its stochastic nature. The algorithm was implemented in LabVIEW, based on the elegant simplified approach of Pedersen and Chipperfield [9].

To compare the performance of the various optimisation approaches, they were applied to the crest factor optimisation of digitally modulated multitone signals of various lengths. The total waveform length was 512 samples in each case. Both the particle swarm optimisation (PSO) and compression algorithms were limited to a run time of 5 seconds. They therefore represent both a comparable computational burden and a convergent state. Since the excitation need only be generated once, the time taken for its optimisation is not an issue in practice.

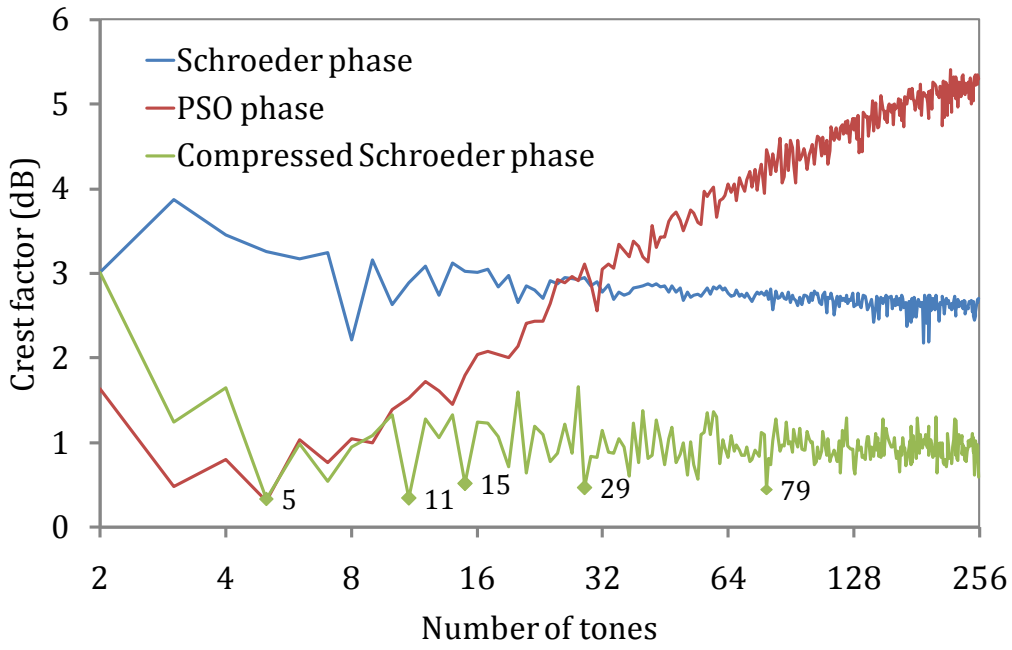



*Figure 5.13: Crest factor for digitally modulated (fs/4) multitone signals with various numbers of tones. Tone numbers with outstanding crest factors are highlighted.*

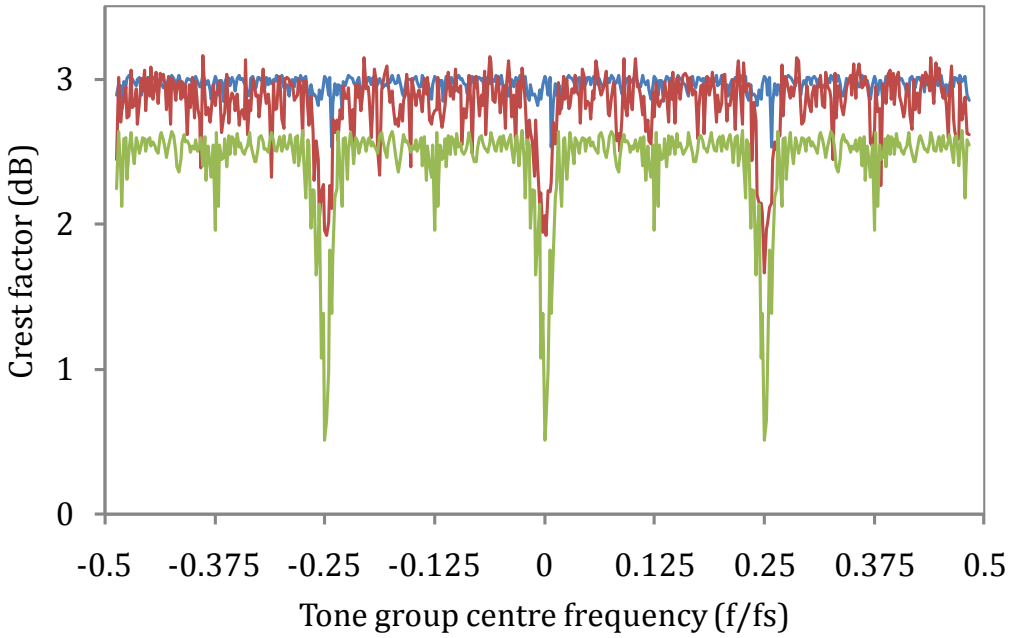



*Figure 5.14: Crest factor for digitally modulated 15-tone multitone signal at various digital modulation frequencies. Key as in Figure 5.13.*

The comparison in Figure 5.13 shows that the 'soft clipping' compression algorithm almost always performed well, whereas the PSO algorithm performance degrades with increasing number of tones. This is because the number of dimensions for the optimisation problem is equal to $P - 1$, meaning that the size of the search space increases $\propto L^P$. This is a problem known as the 'curse of dimensionality' [10], and is a common problem in machine learning. The compression algorithm, on the other hand, does not increase in complexity at all with larger numbers of tones. Whilst other optimisers may fair better, the compression algorithm with Schroeder starting phase performs well enough that there is no need to pursue this line of investigation further.

The results in Figure 5.14 show that DC-centred and $\pm f_S/4$ centred multitone signals have a clear advantage in crest factor. At $f_S/4$ this is because the modulating wave has the simple representation discussed above. This, therefore, has minimum impact on total waveform crest factor, providing yet another good reason to choose this modulation frequency whenever bandwidth requirements allow. Based on the settings and optimisation used here, the achievable crest factors are less than 10 % more than that of a sine wave, resulting in no more than

1 dB reduction in the SNR of the resulting sensor spectral estimates compared to a single sine wave of the same amplitude.

### 5.1.6 Selecting the optimum number of tones

With crest factor optimisation now providing for performance comparable to a single sine wave, the only problem that remains is choosing the right number of tones. This is a non-trivial problem, as having too few tones will affect the eigenvalue estimate accuracy and precision, as shown in the case of the VNA previously in Section 4.5, whereas having too many tones will decrease each tone's amplitude and therefore the SNR of each frequency point of the spectral estimate.

For a flat-spectrum multitone signal of $P$ tones, limited by hardware to a fixed maximum value, the average tone amplitude is given by:

$$A = |X| = \frac{\max|x[n]|}{\sqrt{P} \cdot C_n\sqrt{2}} \tag{5.18}$$

Thus increasing the number of tones decreases their amplitude and thus the SNR of the spectral estimate at that frequency. However, the precision of the eigenvalue estimate should increase the more frequencies are measured. Ideally this would cancel out the $\sqrt{P}$ dependence altogether.

To see how these considerations are reflected in practice, a simple experiment was performed with the second implementation of the multitone system, using the same temperature-stabilised cavity resonant sensor as used in Section 4.6. Although similar to the VNA measurements, in this case the tone spacing was constant at just under 200 $KHz$ for a 512-sample excitation signal. The tones were also centred at $f_S/4$, therefore the effect of adding tones was to increase the effective span rather than the frequency resolution. The results are shown in Figure 5.15. Lorentzian fitting was used to extract eigenvalue estimates. Judging by the apparent convergence of the mean value of the eigenvalues, it seems as if the accuracy does improve with increasing tone number, particularly in the bandwidth measurement, although it is difficult to conclude from this result how significant

this effect is. However, the peak-to-peak deviation (Equation 4.9) does increase linearly with the number of tones. The results seem to suggest that a reasonable compromise is reached with ~ 31 tones, but this will likely be highly specific to the average resonant sensor operating conditions (*e.g.,* average bandwidth) and the number of samples used.

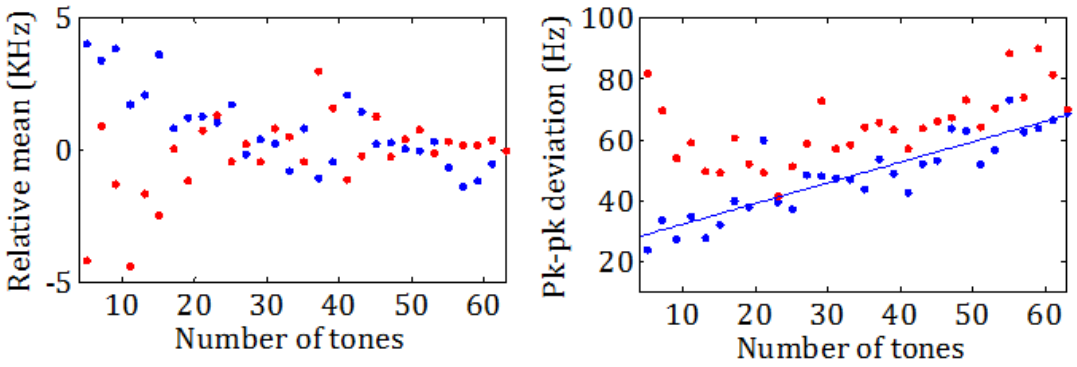



*Figure 5.15: Comparison of the effect of a variable number of tones measured using the multitone hardware system and a temperature-controlled static resonator. Red points represent bandwidth measurements and blue resonant frequency.*

### 5.1.7 Adaptive filters for response estimation

Adaptive filtering based on the stochastic gradient method (*i.e.,* using an iterative weight update rule) is a well-known technique that has been actively researched for many decades, and the classic least mean squares (LMS) algorithm was proposed by Widrow and Hoff in 1960 [11], although the general concept of was proposed even earlier [12]. It is not necessary to present the theoretical background to adaptive filters here, as this is not a digital signal processing focussed thesis, and this application of them is also very straightforward. For background information, excellent references include Haykin, [13] and Sayed, [14].

Applications linked to microwave engineering include communications (*e.g.,* wireless channel equalisation) [15] and radar (*e.g.,* target tracking) [16]. Estimation of the resonant sensor transfer function can be done using a specific

adaptive filtering mode known as parameter tracking. In this mode, the adaptive filter is configured to track changes in the time-varying parameters of a system.

Unlike static filters or FFT-based processing, adaptive filters cover a wide superset of situations where system dynamics and signal statistics are not stationary but vary at any rate up to the Nyquist frequency. MRS parameter tracking is a middle-ground application, lying between that of recovering a noise-corrupted signal on the one hand (where the signal is approximately stationary in a statistical sense) and detecting a sudden fault or change in system dynamics on the other (where the signal can undergo sudden changes of statistics). The strength of the adaptive filter approach, compared to the FFT algorithm discussed above, is this ability to cope well both with very rapid and very slow changes in the resonant sensor. Adaptive filters have not been used before for processing microwave sensors, and their use in this application was incorporated into the patent filed on the multitone measurement system [17]

The adaptive filter is configured to predict the sensor response, $y[n]$ based on the excitation signal $\mathbf{x}[n]$ over the previous $N$ samples. In other words, the adaptive filter tries to model the sensor, producing an estimate of the sensor impulse response $\hat{\mathbf{h}}[n]$ that strives to match the true sensor as closely as possible, from this point of view working in much the same way as curve fitting. Theoretical analysis of the most common adaptive filter algorithms shows that, in the long-time limit, they converge to the best estimate of the true impulse response in the least-squares sense [14], minimising the cost function $C \triangleq \left\|\hat{\mathbf{h}}[n] - \mathbf{h}[n]\right\|^{\mathbf{2}}$. They are also extremely efficient at it, the simplest adaptive filter algorithm (LMS) being computationally more efficient than an averaged FFT approach [18].

Adaptive filters can thus track changes in the sensor spectral response over time. This is demonstrated in the simple simulation results in Figure 5.16. In the top image is a spectrogram shows the time-frequency representation of a microwave resonator. The dark blue band shows a resonance initially centred in the 100 MHz window. Over time, there are three sudden step changes: to a higher frequency, then to a broader bandwidth, and then back to the original resonant frequency. In

the lower image is the shown the FFT of the adaptive filter weights over the same period. The adaptive filter weights can be seen to estimate and track the changes in the resonator response in the presence of noise. Its response to these step changes is similar to a simple low pass filter, and is caused by the filter re-converging after a step change at a finite rate. The characteristic time for this is tuneable through parameters of the adaption algorithm used, and there exists a simple inverse relationship between the settling time and the noise reduction properties of an adaptive filter.

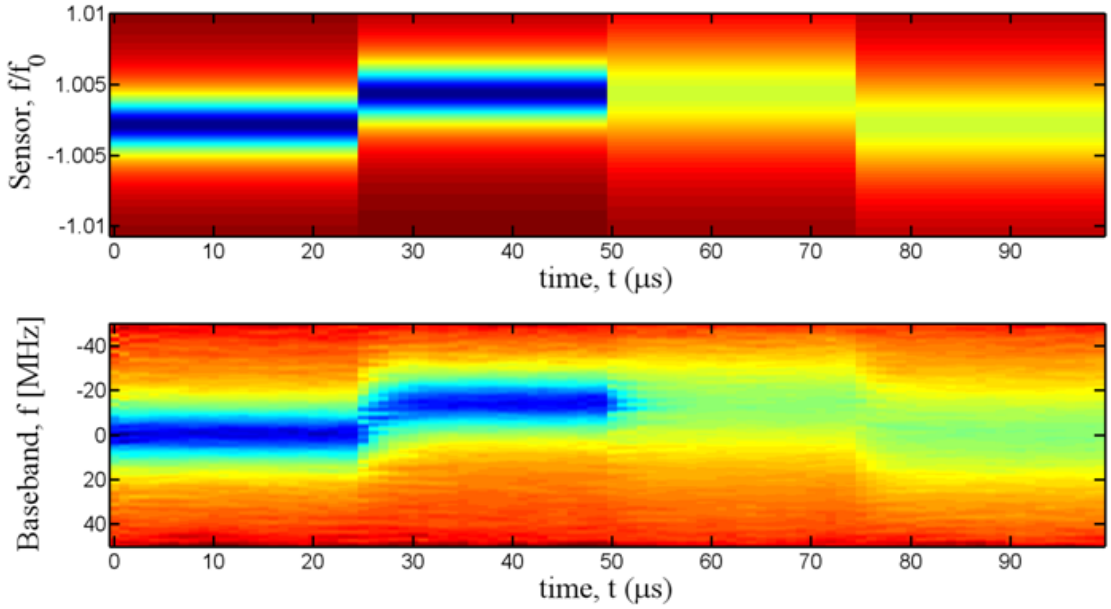



*Figure 5.16: Spectrogram showing an adaptive filter tracking step changes in the resonant frequency and bandwidth of a simulated microwave resonator. The adaptive filter spectrogram (bottom) was obtained by calculating the Fourier transform of the adaptive filter weight vector.*

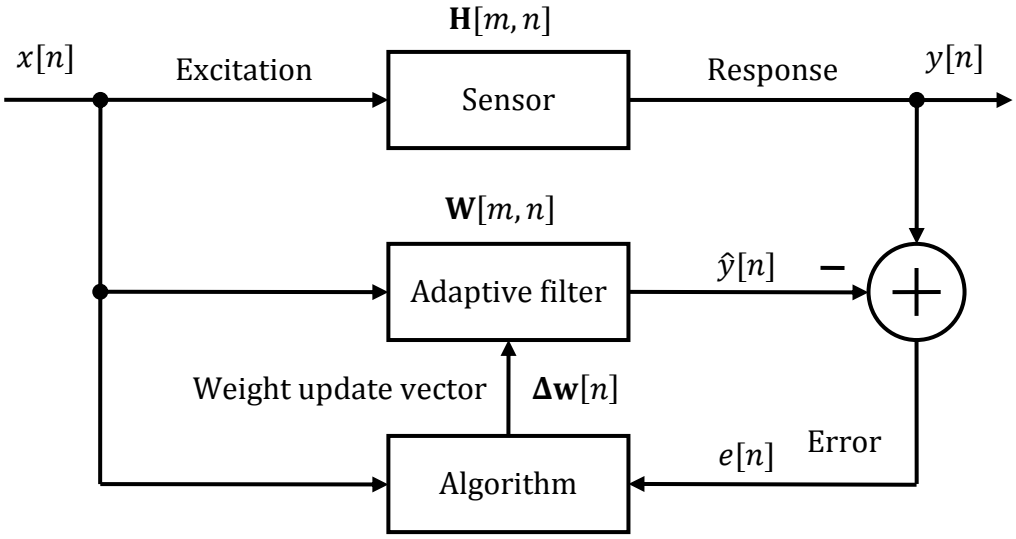



*Figure 5.17: General adaptive filter configured for system identification.*

There are two types of digital filter, infinite impulse response (IIR) and finite impulse response (FIR). IIR filters use both feed-forward and feedback, in general, giving a transfer function with arbitrary coefficients on both the numerator and denominator. FIR filters, on the other hand, only implement a simple feed-forward scheme, equivalent to the dot product of the filter weights and a vector of time-delayed samples. They thus have only finite-length effect on the signal samples.

Whilst IIR adaptive filters do exist, their convergence is tricky as they have stability issues. A second-order IIR adaptive filter is effectively a tracking pole. It could therefore be used to track the resonant sensor transfer function efficiently with only three adaptation weights. In theory, the sensor eigenvalue would also be simply related to this IIR filter pole, completing the eigenvalue estimation task in an elegant way. However, in practice this does not work because the digitally sampled sensor resonance can never truly match a digital IIR filter resonance due to the effect of aliasing. Thus, this method suffers from poor accuracy that makes it impractical for an MRS system.

Thus the rather less elegant but much higher performance method is to implement a simple and robust FIR adaptive filter, with the same length as the excitation waveform, and use its Fourier-transformed weights as a fast, time-domain estimate of the sensor transfer function, updated as they are every time sample.

Depending on the adaptive filter algorithm, given a step-change in the sensor transfer function, the coefficients of the filter will converge on the impulse response samples exponentially in time or faster. The filter algorithm thus provides some controllable low-pass filtering functionality. Proper selection of the adaptation algorithm parameters should ensure that the signal to noise ratio of this spectral estimate is roughly equal to the signal to noise ratio of the input signals. The adaptive filter will contribute its own 'estimation' noise, due to the stochastic nature of the algorithm itself. Careful control of the algorithm parameters is therefore needed to balance the convergence rate with the suppression of this estimation noise.

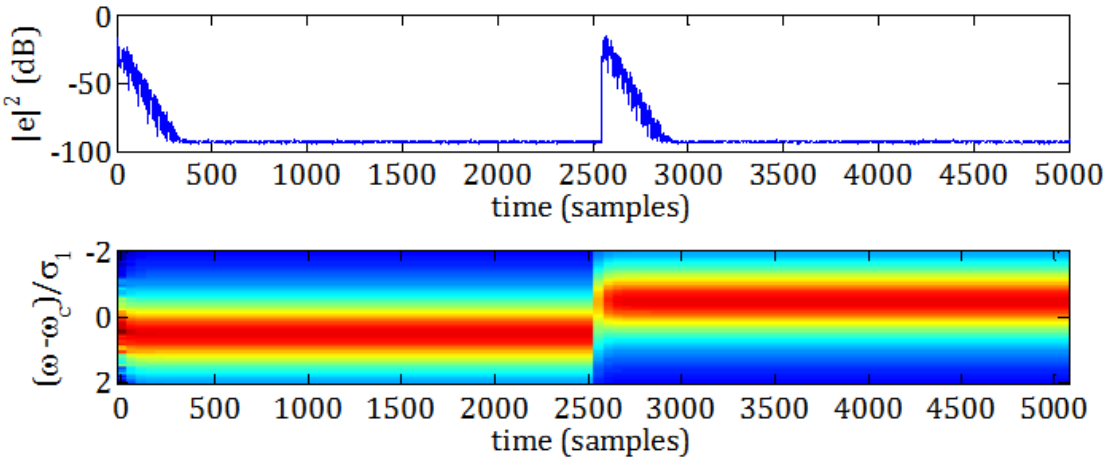



*Figure 5.18: Simulation of the NLMS filter algorithm applied to resonator transfer function estimation in the time domain. The top figure is the time-domain prediction error in dB - the input noise level was 80 dB. A sudden jump in resonant frequency is present in the middle of the record. This is reflected in the readjustment error and the spectrogram of* $\hat{\boldsymbol{H}}$ *in the bottom figure.*

Assuming an FIR adaptive filter model, the equations for a general stochastic-gradient adaptive filter are:

| True sensor equation | Adaptive filter estimate |
|---|---|
| $y[n] = h[n] * x[n]$ | $\hat{y}[n] = \mathbf{w}^H[n-1] \cdot \mathbf{x}[n]$ |
| Weight update equation | Weight adaptation algorithm |
| $\mathbf{w}[n] = \mathbf{w}[n-1] + \mathbf{\Delta w}[n]$ | $\mathbf{\Delta w}[n] = \mu[n] \cdot \mathbf{f}(\mathbf{x}[n]) \cdot e[n]$ |

5.19

The step size $\mu[n]$ and gain $\mathbf{f}(\mathbf{x}[n])$ are specific to the algorithm. Whilst many sophisticated adaptive filter algorithms exist, only very simple algorithms have any real place in this application, as the number of computations required per second would otherwise be impractically high. The two algorithms investigated were the normalised least mean squares (NLMS) algorithm [14] and a change detection algorithm [18].

The following equation describes the (well-known) algorithm of the NLMS filter used in this 'system identification' mode:

$$\hat{\mathbf{w}}[n] = \hat{\mathbf{w}}[n-1] + \mu \frac{\mathbf{x}^*[n]}{\mathbf{x}^H[n]\mathbf{x}[n]} (y[n] - \hat{\mathbf{w}}^H[n-1]\mathbf{x}[n]) \quad 5.20$$

Where $\hat{\mathbf{w}}[n]$ are the filter weights at the $n$th iteration, $\mu$ is the scalar constant learning rate, $\mathbf{x}[n]$ are the $N$ excitation samples, $y[n]$ is the last received sample, and $(.)^H$ denotes the Hermitian transpose of a complex vector. The expression $(y[n] - \hat{\mathbf{w}}^H[n-1]\mathbf{x}[n]) = e[n]$ is equivalent to the error between the true output sample at time $n$ and that predicted by the adaptive filter. Often in definitions of this algorithm, a small constant is added to the denominator to eliminate divide-by-near-zero errors. However, in this case $\mathbf{x}[n]$ is known to be persistently exciting and therefore this is not necessary.

In this baseband implementation, the weights vector, received signal, and excitation vector are all complex. $N$ is the filter length. The spectral estimate of the resonator is gained by the Fourier transformation; $\hat{\mathbf{H}}[n] = \mathcal{F}\{\hat{\mathbf{w}}[n]\}$. Because the excitation is fixed and periodic within $N$ samples, it is logical to implement the above adaptive filter in block form, with length $N$. The term, $\mathbf{x}^*[n](\mathbf{x}^H[n]\mathbf{x}[n])^{-1}$, can then be pre-computed as a $N \times N$ Toeplitz matrix $\mathbf{X}_n$ and stored in memory.

This makes the NLMS algorithm here no more computationally complex than the basic LMS algorithm it replaces, becoming:

$$\widehat{\mathbf{W}}[Nn] = \widehat{\mathbf{W}}[Nn - N] + \mu \cdot \mathbf{X}_n \mathbf{e}[Nn], \tag{5.21}$$

where $\mathbf{e}$ is a vector of the error signal over the previous $N$ samples. This iteration produces an $N \times N$ block of weights $\widehat{\mathbf{W}}$ every $N$ samples. The computational complexity of this algorithm per spectral estimate is $1 + 2N$ multiplications and $3N$ additions, resulting in a total of $2561 \times 100M = 256\ \mathrm{GOP}s^{-1}$ if processed at full rate. This is very high, but not beyond the capabilities of modern FPGAs. Evidently, for this algorithm to have any comutational advantage over the FFT algorithm (complexity order $N\log_2 N$), it isn't practical to FFT the weight vector at this rate. Thus, the weight vector can either be filtered and decimated in the time domain or an eigenvalue extraction method that works directly on the time-domain weights can be employed. The neural network-based eigenvalue method is ideal for this latter task, discussed in the next section.

### 5.1.8 Full system tests with adaptive filter algorithm

In order to assess the performance of the adaptive filter algorithm in estimating the sensor transfer function, the NLMS algorithm was implemented in LabVIEW and used with the first implementation of the multitone hardware (Section 4.4.1). To compare the combined algorithm and system accuracy in estimating the transfer function, a simple experiment was performed.

Using two circulators, a VNA (E5071B, Agilent) and the prototype multitone system were connected simultaneously to the same resonator (Figure 5.19); the same rectangular cavity resonator used in the heating experiments (Section 6.3). The excitation signals interfere, even with the circulators, so central LabVIEW PC control was necessary to switch between the measurement systems smoothly whist still allowing reasonably high sampling rate and realtime performance. This was achieved by disabling the RF output of each system in turn using inbuilt solid-state switches. Acquisitions were also synchronised to this switching.

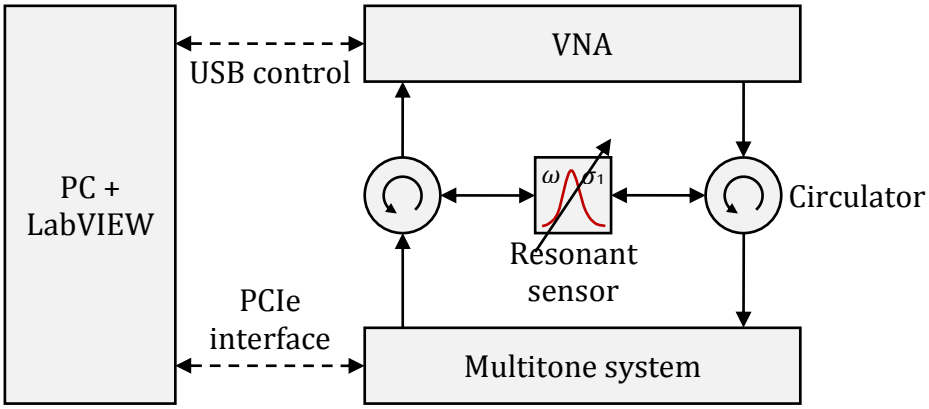



*Figure 5.19: Multiplexing of the VNA and multitone system.*

Example results are shown in Figure 5.20 and Figure 5.21. The VNA was configured with a high IFBW (50 KHz), as a relatively large number of sweep points (215) were used for spectral estimate comparison reasons. The trace noise present is rather higher than it could be if optimised, therefore. However, the multitone system with adaptive filter algorithm is evidently a comparable, if not superior estimate of the spectral response, appearing to have less trace noise and being in good agreement generally with the VNA. The only apparent inaccuracy is in phase, where the deviation becomes significant towards the edges of the FFT window. This is caused by the lack of phase calibration; the response does not take into account the phase of the antialiasing filter roll-off. This could be rectified using static equalisation with a 'thru' calibration standard in place of the sensor.

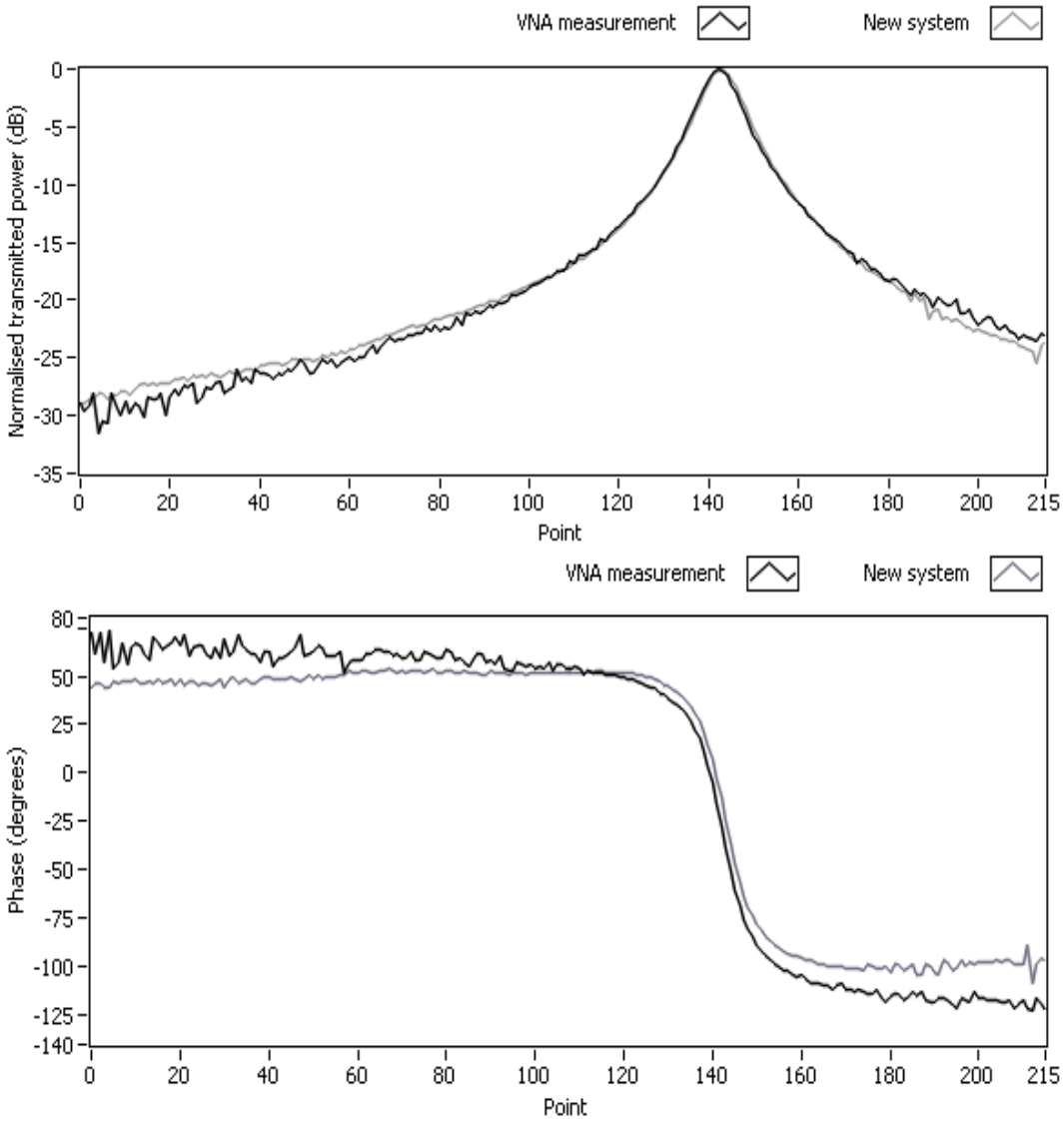



*Figure 5.20: Example of a spectral response estimate with the first prototype system compared with a simultaneous VNA measurement*

When both these spectral measurements are converted to eigenvalue estimate using the Lorentzian curve fitting method, the time-domain tracking of this sensor also shows superiority on the side of the multitone system. The VNA measurements are at a slower mean rate, with around 60 times fewer samples taken over the 5-minute interval (the average rate was 3 ms for the multitone system and 180 ms for the VNA – although the sampling rates were not uniform due to need to switch from one instrument to the other). A matching linear trend can be seen in the resonant frequency measurements, due to temperature drift, and a small offset of about 2 KHz. There is no way of knowing which system is

more accurate, and at this magnitude could be due to the different local oscillators in each instrument, which were not locked together. Assuming this offset is static, it would not cause any error in differential sensor measurements.

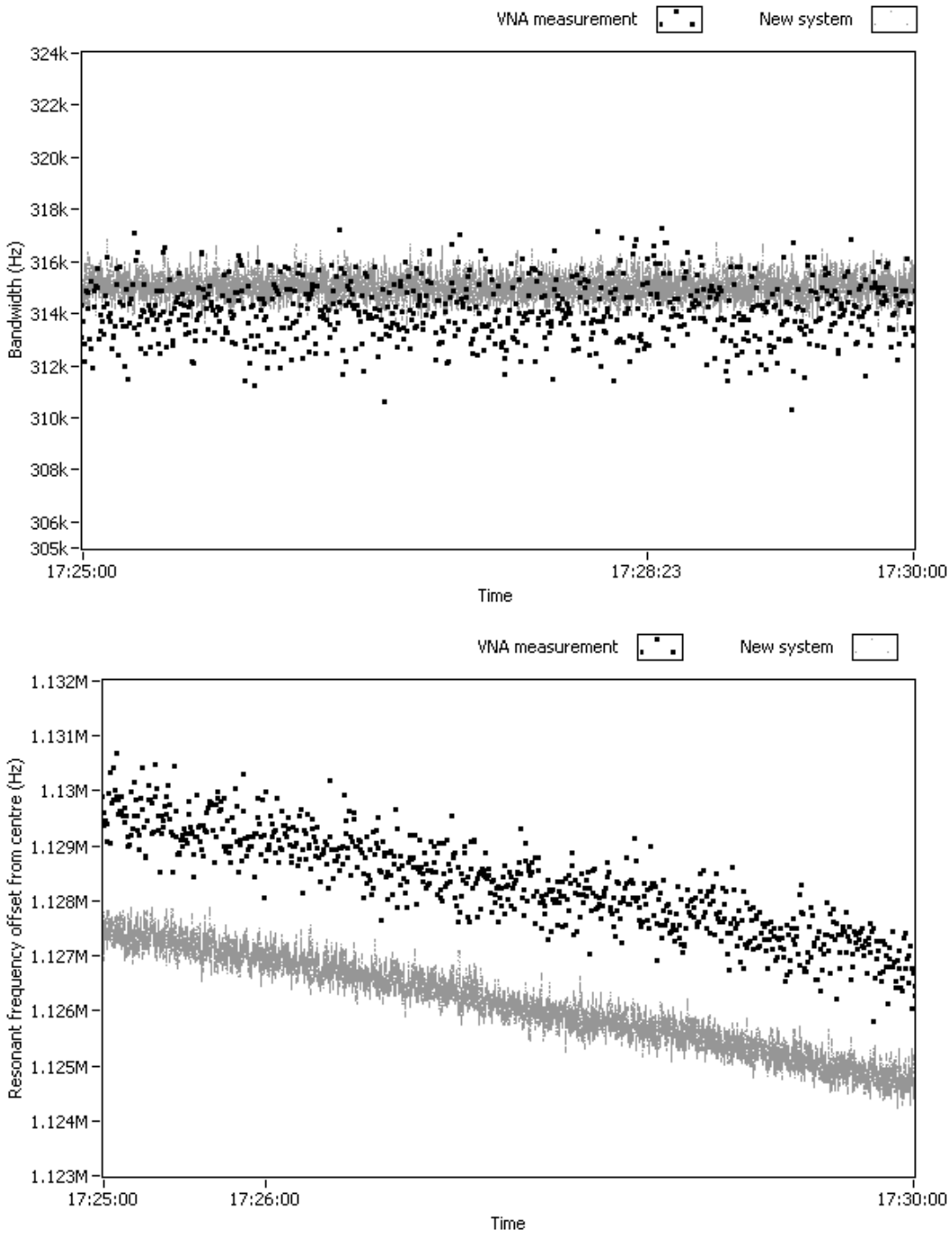



*Figure 5.21: Simultaneous comparison of the first prototype system and VNA during measurement of a static resonator*

## 5.2 Correcting for system imperfections

As discussed above, the estimation of the sensor spectral response – upon which all further processing stages rest – is reliant on the assumptions of linearity and ideal operation in the various stages present in the transmitter-receiver signal chain. This section addresses software methods used to minimise the nonlinearity and distortion of the spectral estimate caused by various hardware imperfections.

As mentioned in the Chapter 4, the most significant cause of distortion is the mixer at the transmitter side. This is because the amount of distortion is dependent on the drive level of the baseband input to the mixer, which is relatively high at the transmitter. In contrast, the RF input to the mixer at the receiver is comparatively less as it has been attenuated by the sensor insertion loss and filtering effect.

### 5.2.1 Nonlinear predistortion for transmitter compensation

Correcting for the distortion using software methods allows the drive level to the modulation mixer to be increased without risking poor performance. This then increases dynamic range and improves SNR without affecting error. This ability to make more efficient use of signal amplitude is why nonlinear predistortion is an essential part of most modern communications systems employing a power amplifier [19], [20].

The nonlinearity of baseband distortion caused by the IQ modulation can be viewed as a two dimensional function of the amplitudes of the ideal in-phase and quadrature signals $i(t), q(t)$:

$$[i'(t), q'(t)] = f_{dist}\big(i(t), q(t)\big), \tag{5.22}$$

where $i'(t), q'(t)$ are the distorted outputs. The concept of digital predistortion is to apply a function to the discrete versions of these signals before the are generated, so that

$$[i'(t), q'(t)] = f_{dist}\big( \quad f_{pre}\big(i(t), q(t)\big) \quad \big) \approx [i(t), q(t)]. \qquad 5.23$$

Thus, the digital prediction attempts to implement $\hat{f}_{dist}^{-1}$ and hence minimise the cost function $C = \|[i'(t), q'(t)] - [i(t), q(t)]\|^2$. Whilst many clever adaptive methods exist to do this, *e.g.* [21], in the multitone system the nonlinearity is largely static in time and, therefore, can be corrected for by a simple fixed predistortion applied to the excitation waveform before it is loaded into the memory of the AWG.

In order to find this static estimate of $\hat{f}_{dist}^{-1}$, the multitone system is calibrated with a known training waveform. The received signal can then be compared to this training waveform and the error function turned into a predistortion look-up-table. Due to IQ gain and phase offset errors in the quadrature modulator, the distortion function is a two-dimensional function, in general. Thus $i'(t)$ is a nonlinear function of both $i(t)$ and $q(t)$. This is a form of nonlinear mixing between these ideally orthogonal signals, meaning that they can no longer be treated as an approximation to a complex number and must be treated as general interdependent variables. Note, however, that gain and phase offset – the 'first order' mixing effects are not an issue *per se,* due to the choice of tone location, as discussed in Section 4.3. Even if it is an issue, this is best dealt with by a separate method, *e.g.,* [22], [23], as it tends to be highly frequency-dependent.

To find this 2D function, the training waveform is derived from a randomly permuted grid-sampled amplitude signal, equivalent to high-density QAM signal (Figure 5.22). This makes sure that the amplitude space is fully sampled, but that an approximately random signal is still created in the time domain. The random nature of the permuted training waveform eliminates sensitivity to linear frequency response, as each random permutation has a slightly different frequency content, and thus over a number of averages with different permutations, only the frequency-independent part of the nonlinear distortion is learnt. In effect, this permutation method extracts the memoryless part of the IQ mixer nonlinearity – this is all that is required, as the tone-flattening algorithm (Subsection 5.2.2) will take care of any remaining frequency dispersion.

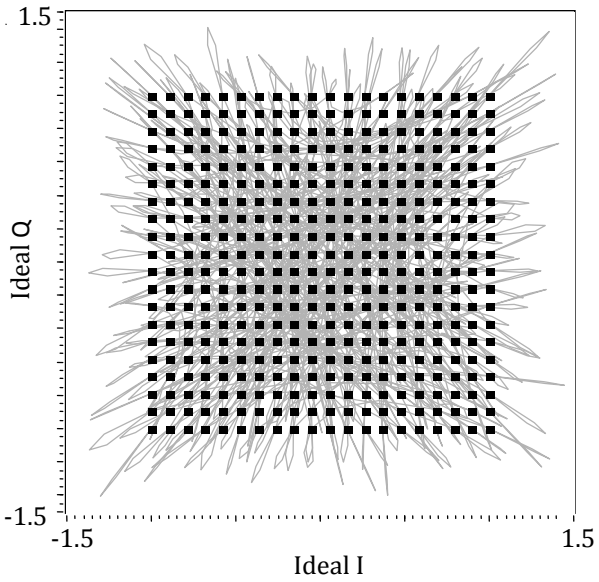



*Figure 5.22: Regularly sampled IQ calibration constellation shown at the data rate (black squares) and the trajectory of the resulting calibration signal after interpolation and root-raised cosine filtering (grey curve).*

To derive the 2D nonlinear predistortion look-up data, the sensor is first replaced with a 'thru' – in this case a spectrally flat variable attenuator set to the same insertion loss as the unperturbed sensor. The training waveform is then generated before passing through an interpolating root-raised cosine filter. An interpolation factor of 4-8 was found to be suitable. This pulse-shaping filter, matched at the receiver, avoids aliasing by bandlimiting the training waveform without affecting the amplitudes at the time samples corresponding to the sampled amplitude grid.

Figure 5.23 shows the measured training waveform at the receiver, before and after predistortion correction. The distorted IQ grid has been largely restored by predistortion at the transmitter.

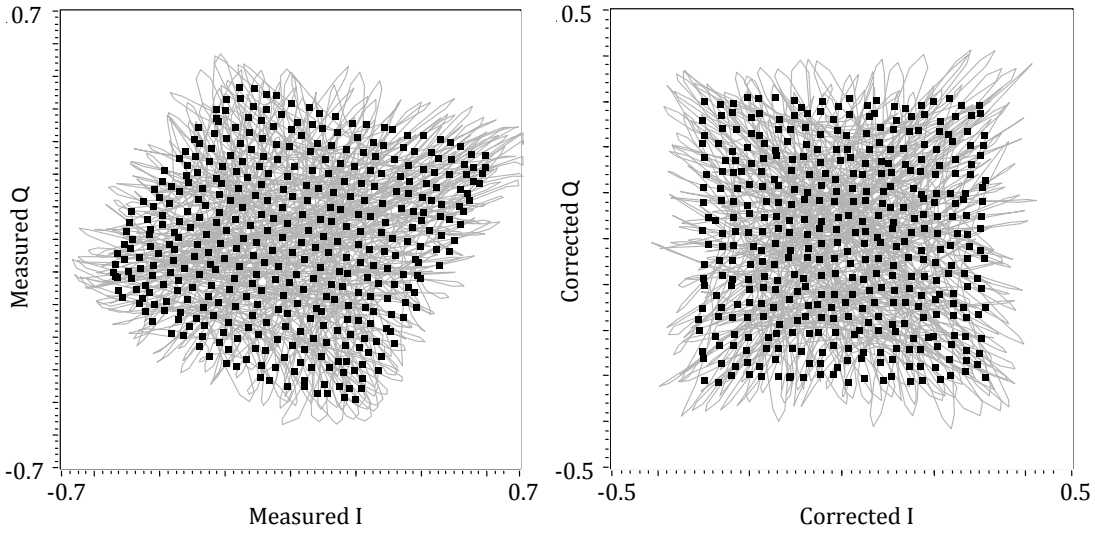



*Figure 5.23: Received IQ calibration constellation before (left) and after (right) predistortion correction.*

In order to apply the predistortion function to an arbitrary excitation signal, the correction at each one of these discrete amplitude sample points needs to be interpolated. Figure 5.24 shows an example of the learnt inverse function $\hat{f}_{dist}^{-1}$, interpolated using a radial basis function (RBF) network, as described in Section 3.6.1. This implements the predistortion estimate function

$$\begin{bmatrix} i_p[n] \\ q_p[n] \end{bmatrix} = \sum_{k=0}^{n} \begin{bmatrix} w_i[k] \\ w_q[k] \end{bmatrix} \phi\left( \begin{bmatrix} i[n] \\ q[n] \end{bmatrix}^2 - \begin{bmatrix} i_0[k] \\ q_0[k] \end{bmatrix}^2 \right), \tag{5.24}$$

where the weights $\mathbf{w}_i$ and $\mathbf{w}_q$ are found by solving the linear system of equations derived from the training waveform and its measured response. The inverse quadratic radial function was found to perform best, $\phi(r) = 1/[1 + (\beta_k r)^2]$, and the centres $\mathbf{i}_0$ and $\mathbf{q}_0$ were chosen from another regularly sampled grid, this time over $I'Q'$ space. A constant $\beta_k = 0.1$ was chosen, resulting in a very smooth interpolation that performed well in practice. The number of basis functions was chosen to be $16 \times 16$, giving a good balance between predistortion accuracy and the amount of training data required to give a good estimate of the RBF weights.

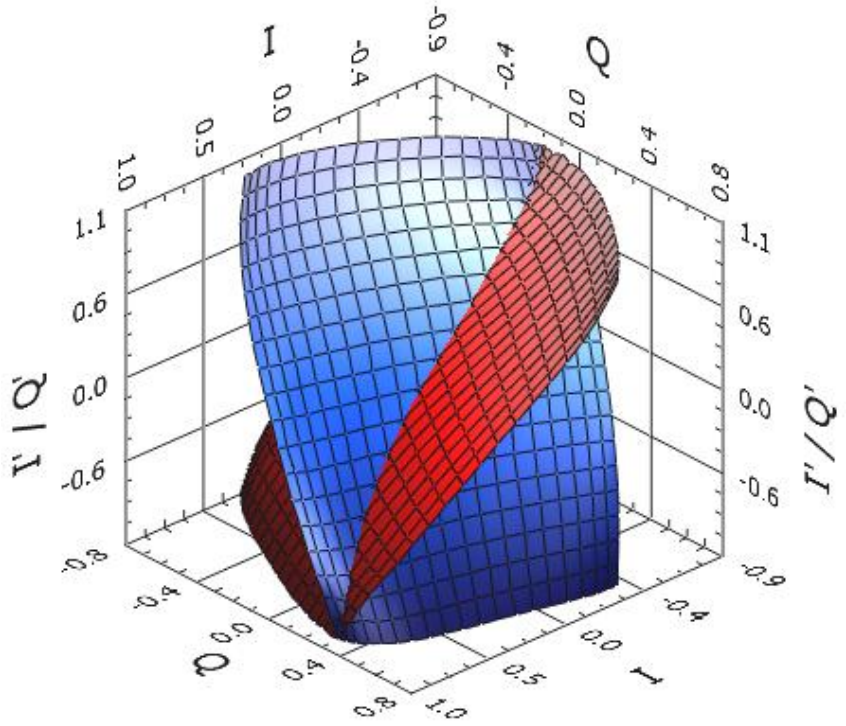



*Figure 5.24: Example of the 2D predistortion function learned by the RBF network. The blue surface represents the mapping between the input IQ plane and the predistorted I' signal. The red surface represents the same mapping for the Q' signal.*

Figure 5.24 demonstrates the impact of the RBF predistortion function on a simple two-tone signal generated by the later implementation of the multitone hardware system. The spectrum was generated by the receiver. The main problem for modulated multitone signals is the in-band third order IMD, which in this case has been reduced by over 10 dB, down almost to the noise floor. Apart from the inconsequential IQ offset image tones reflected about DC (bin 32), the other distortion products have all been reduced, giving over 50 dB of dynamic range, without altering the amplitude of the two-tone signal significantly. The predistortion algorithm was found to be very important for satisfactory performance of the full-bandwidth measurements on the split-ring sensor, discussed in the Section 6.2.

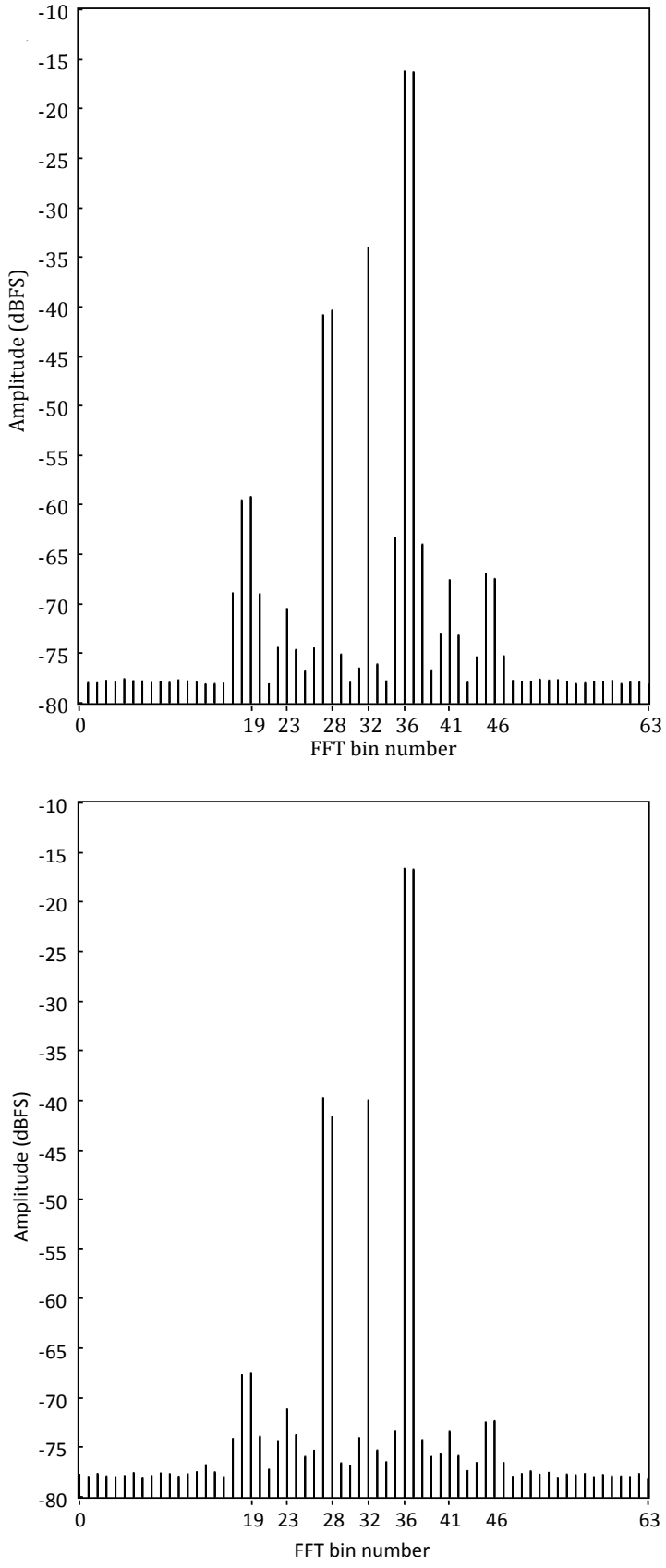



*Figure 5.25: Two-tone signal comparison before (top) and after (bottom) predistortion using the RBF interpolating function.*

### 5.2.2 Tone flattening algorithm

During a calibration phase, a further correction that can be performed is to correct for small variations in the measured multitone amplitudes as measured at the receiver. Flattening the received tone in this way not only compensates for the roll-off and ripple of the antialiasing filters, *etc.,* it also helps reduce the impact of distortion at the modulator even further. Any residual, static variation in the relative tone amplitudes caused by nonlinear intermodulation can be compensated for by slightly increasing or decreasing the amplitudes of the tones transmitted.

Due to the effect of system nonlinearities, the compensating predistortion algorithm, and the need re- optimise the crest factor after changing its tone amplitudes, the relationship between the amplitudes of the generated tones and the received tones is complicated. Changing the amplitude of just one tone has a small yet finite effect on the amplitude of all the others. This relationship is unknown, and it is impractical to estimate it accurately, even to first order.

Consider a multitone signal with 32 tones. A general first order relationship between the generated and received tones could be represented as $\mathbf{y} = \mathbf{M}\mathbf{x}$ where $\mathbf{y}$ is a vector of received tone amplitudes and $\mathbf{x}$ is a vector of generated tone amplitudes. The 'off-diagonal' elements of this matrix account for nonlinear effects. $\mathbf{M}$ is a matrix with, in general, 1024 distinct coefficients. To sample this 32-dimension space fully, in order to estimate this matrix reliably, the simplest full-factorial experiment would require at least $2^{32} \cong 4.3$ billion experiments, due to the very high number of degrees of freedom of this matrix. Estimation of the interrelation between generated and received tone amplitudes as a route to direct multitone flatness calibration is clearly impractical, even if the number of coefficients can be reduced by only considering nonlinear interactions up to a certain order.

Therefore, a simple iterative method was adopted instead. The 'hardware-in-the-loop' iterative tone-flattening algorithm developed to solve this issue is necessarily the last step in the calibration process. It ideally compensates for the attenuation

caused by linear dispersion effects of the RF system, reconstruction, and antialiasing filters once the system has been linearised by the predistortion algorithm, as well as 'mopping up' the effects of small nonlinear mixing occurring at the transmitter side.

Using the averaged-FFT algorithm to estimate the relative magnitude of the received tones, $|\mathbf{Y}[i]|$, an iterative tone amplitude correction algorithm was developed, that determines a new value for the tone amplitude vector $\mathbf{A}[i+1]$ based on the relative deviation of the received tone amplitudes from flat, *i.e.,* minimising the cost function $C = \| \, |\mathbf{Y}[i]|^2 - \text{median}|\mathbf{Y}[i]|^2 \|$:

$$\mathbf{A}[i+1] = \exp[\, \ln \mathbf{A}[i] - \gamma \ln(\, |\mathbf{Y}[i]|^2 - \text{median}|\mathbf{Y}[i]|^2 \,)]. \qquad 5.25$$

The median tone amplitude is used, rather than the mean, as the relative tone amplitudes are not normally distributed in general, thus this gives a better estimate of the relative amplitude deviation. At each step, crest factor optimisation on the resulting multitone waveform was performed using the soft clipping algorithm. A logarithmic version of the algorithm was found to perform better than a linear version – in effect performing adaptation in the logarithmic domain. The adaptation step-size parameter $\gamma$ was chosen to be 0.1 to give a slow but reliable convergence to flatness.

The typical convergence of this iterative algorithm is shown in Figure 5.26, showing that convergence is reached within about 20 iterations to a peak-to-peak deviation of just 0.02 dB. Figure 5.27 shows the effect of the tone-flattening algorithm on the measured multitone spectrum at the receiver for a wideband excitation signal (100 MHz total span). In this case, the peak-to-peak deviation is around 0.03 dB. Whilst the algorithm cannot be applied in an online fashion, as the need for crest factor optimisation makes it quite slow, the effect of drift on the tone amplitudes should be minimal if the system is temperature controlled. Thus, the tone amplitude flatness can be expected to be improved significantly even from this one-off correction.

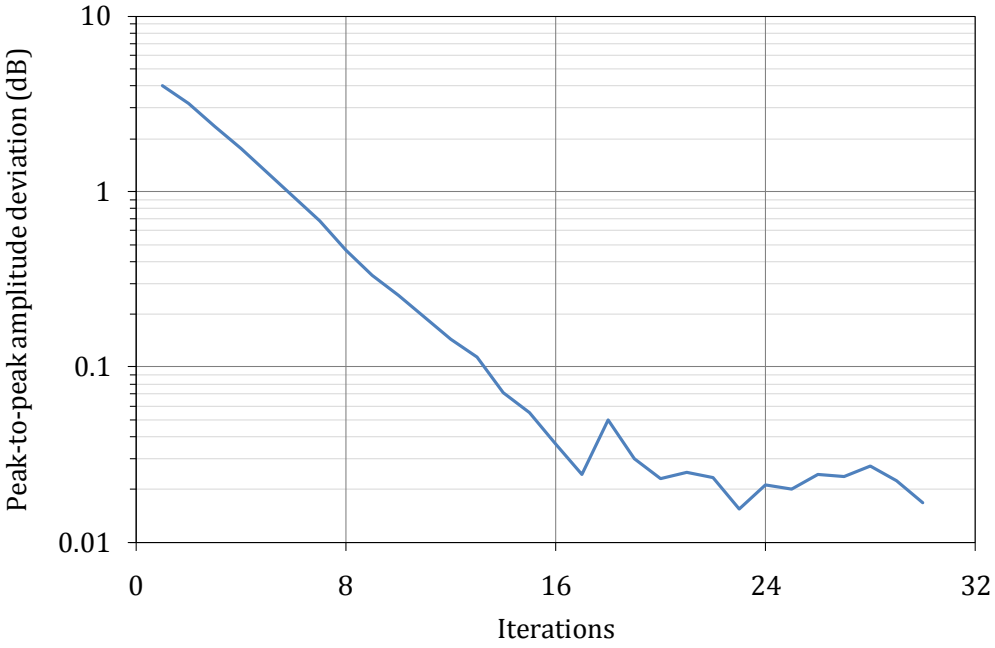



*Figure 5.26: Typical convergence performance of the multitone flattening algorithm with number of iterations.*

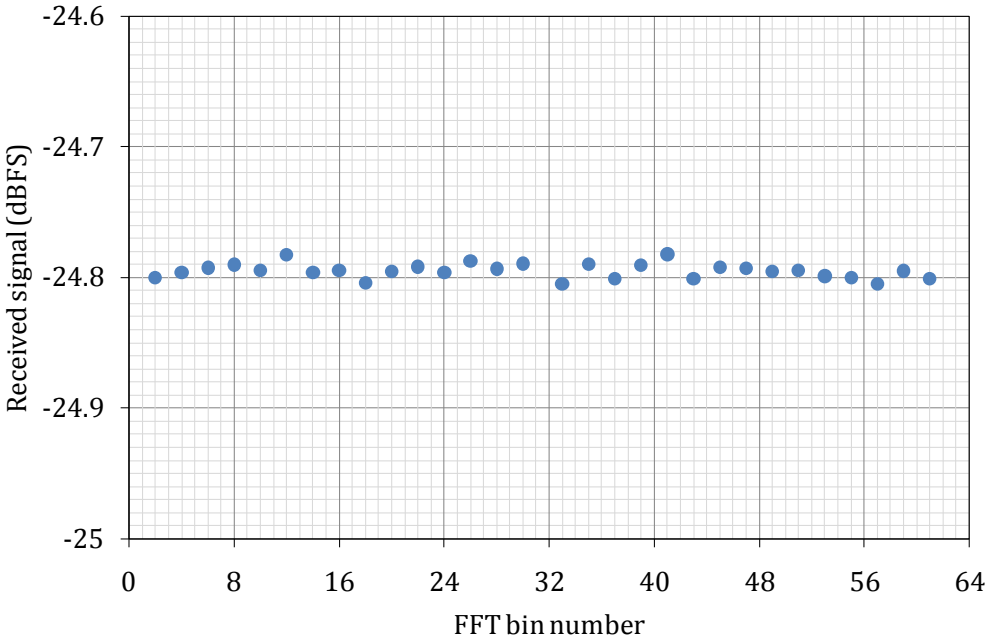



*Figure 5.27: Measured amplitude of multitone signal at receiver during calibration phase after tone flattening optimisation algorithm was run for 30 iterations. Total peak-to-peak deviation from flat is less than 0.03 dB.*

## 5.3 Inversion of the spectral response function

In the Theory chapter (Subsection 3.1.2), the multistage model was introduced. This section discusses software methods to invert Stage 3 of this model – the link between (time-varying) eigenvalue and (time-varying) spectral response:

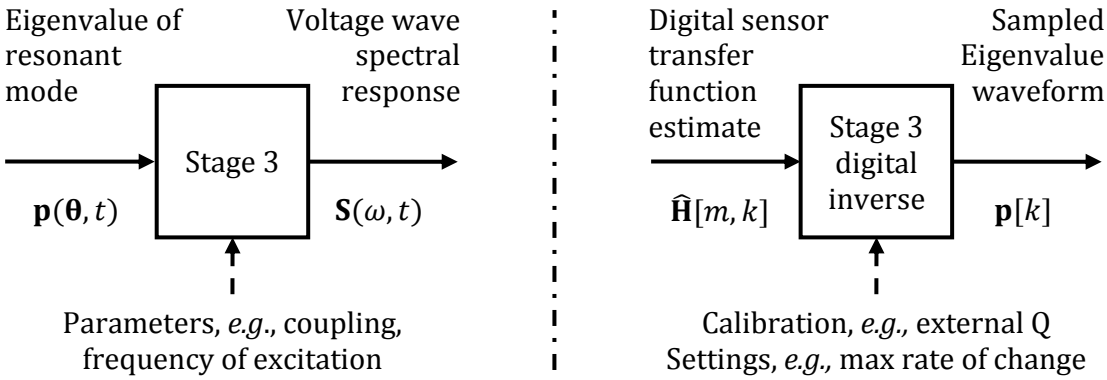



*Figure 5.28: Stage 3 of the resonant sensor model (left) and the inverse function to be implemented in the digital software (right)*

This stage generally means extracting parameters for some spectral response model as a function of time, *e.g.,* in the notation I have adopted:

$$S_{ij}(\omega, t) = K_0 + \frac{K_1 \omega_1(t)/2}{\sigma_1(t) + j\big(\omega - \omega_1(t)\big)}, \qquad \mathbf{p}(t) = \begin{bmatrix} 1 & 0 \\ (2Q_e)^{-1} & 1 \end{bmatrix} \begin{bmatrix} \omega_1(t) \\ \sigma_1(t) \end{bmatrix}, \qquad 5.26$$

Historically, this process has received the most attention in the literature, as it has always been a necessary part of resonator measurements regardless of the method used to estimate their spectral response. Almost all of the previous work in this area has been related to VNA-based resonator measurements, and particularly to the measurement of unloaded Q factor. This is an important diagnostic measurement used for quality control of microwave substrate materials and dielectric resonators. The first methods aimed at analogue network analysers were graphical, based around Smith chart measurements [24] and the measurement of specific points around the resulting resonance circle, after Ginzton in 1957 [25].

Circle fitting remains a powerful estimation method [26], [27], often used as a 'first' step to correct for the factors $K_0$ and $K_1$ as well as other imperfections.

The authority on Q factor measurement is Darko Kajfež, whose books [28],[29] and papers[24], [30],[31] develop a rigorous method of estimating Q factor from circle fitting in the complex plane. This method implicitly generates fit parameters that can be converted into bandwidth and resonant frequency, as well as Q. However, as is evident from the Theory section, Q factor is not a very useful parameter for MRS measurements as it does not relate linearly (even approximately) to measurands like complex permittivity. Thus, the high accuracy of Q factor measurement methods (such as Kajfež's QZERO software [32]) does not necessarily imply that these methods will give a high measurement accuracy overall. Furthermore, they are based on the assumption of measurement on a calibrated vector network analyser. Whilst these techniques may be suitable for accurate estimation of the $Q_e$ constant of a given sensor (for calibration purposes), they are not very suitable for use in an online system.

Pertersan and Anlage [33] published a useful comparison of curve-fitting based methods, with both precision and accuracy determined by both numerical simulation and from measurements of a high-Q superconducting cavity resonator. They also discuss methods that can correct for power-dependent quality factor nonlinearity, something that can occur in superconducting resonators (for example in type II superconductors when localised microwave current densities exceed a critical value causing localised transitions into a higher resistance mixed conduction state). They found that the Lorentzian fit (actually, a skewed Lorentzian fit) performed very well, and was the best overall performer over a wide range of input signal to noise ratios. Full complex-domain regression is still superior, however, as shown in [34]. Other methods consist of a sequence of various curve fits, such as circle fitting to correct for rotation and offset, followed by fitting to the phase response. For superconducting MKID resonators, Gao developed a multistep curve fitting approach [35]. In this thesis, I have used a combination of fitting methods, including Lorentzian and full complex domain.

### 5.3.1 Online parameter extraction via curve fitting

The vast majority of methods are therefore based on curve fitting. Curve fitting is a mathematical procedure in the field of regression, whereby the parameters (coefficients) of a mathematical model that best agree, by some measure, with collected data are found using an algorithm. Usually, the measure of 'agreement' of a mathematical model's prediction with data is the squared error, $\mathbf{e} = (\hat{\mathbf{y}}(\mathbf{a}) - \mathbf{y})^2$, where $\mathbf{y}$ is a vector of data points and $\hat{\mathbf{y}}$ are the model predictions based on the fitted parameters, $\mathbf{a}$. Curve fitting is therefore an optimisation process, minimising a sum-squared-error cost function $C = \|\mathbf{e}\|^2$ as a function of the parameters, $\mathbf{a}$

When the function $\hat{\mathbf{y}}(\mathbf{a})$ is nonlinear, as in the case of resonant sensor response, the optimisation becomes tricky, as the relationship between the gradient of the cost function, which sets the best 'direction' for its minimisation, $-\boldsymbol{\nabla}_{\boldsymbol{a}} C$, is also nonlinear. The optimisation may thus suffer from a host of problems, such as multiple minima (*i.e.,* multiple solutions to $\boldsymbol{\nabla}_{\boldsymbol{a}} C = \mathbf{0}$), difficulty with convergence (*e.g.,* taking a long time, or not converging at all, or diverging), as well as sensitivity to the initial starting point for the fit algorithm (the value of the first iteration $\mathbf{a}[0]$) [36]. The link between the statistics of the fitted parameters (*i.e.,* variance and bias), and the statistics of the collected data points is also nonlinear and can be impossible to predict through any other means than Monte Carlo simulation.

Whilst the superiority of curve fitting for eigenvalue estimation is not in question, one of the issues addressed in this thesis was how to perform curve-fitting best when the sensor is part of a realtime measurement system. Thus, issues such as speed, computational complexity, reliability, and robustness become important. If transfer function estimates are generated at KHz or even 100s of KHz, the task of extracting an eigenvalue estimate from each one by curve fitting becomes extremely challenging. This is because nonlinear curve fitting is computationally intensive, requiring many iterations, each consisting of several matrix operations. Due to these special requirements of high-speed MRS readout, general-purpose curve fitting algorithms as implemented in MATLAB, LabVIEW, etc., were not up to the task. Optimised algorithms were thus developed that can run at a much higher

rate and include specialisations that provide increased robustness for unsupervised, online MRS systems.

In the same way that the adaptive filter algorithm is an iterative, stochastic approach to the minimisation of a cost function, iterative curve fitting algorithms can be 'unwrapped' in the time domain. Thus, in contrast to the traditional method of performing many time-consuming iterations on the same data, each iteration can be done with fresh data, generating a new parameter estimate. This is ideal for an online system, as the previous iteration provides the initial starting point for the next. When unwrapped into a time-domain recursive algorithm like this, the classic nonlinear least squared curve fitting technique, the Gauss-Newton algorithm, becomes the second-order stochastic gradient descent algorithm (NLMS, by comparison, is a first-order stochastic gradient descent algorithm) [13]:

$$\begin{aligned} \hat{\mathbf{a}}[n+1] &= \hat{\mathbf{a}}[n] - \mu \frac{\boldsymbol{\nabla}_{\mathbf{a}} C(\hat{\boldsymbol{a}}[n])}{\boldsymbol{\nabla}_{\mathbf{a}}^2 C(\hat{\boldsymbol{a}}[n])}, \quad or \\ \hat{\mathbf{a}}[n+1] &= \hat{\mathbf{a}}[n] - \mu\, (\mathbf{J}^T \mathbf{W} \mathbf{J})^{-1} \mathbf{J}^T \mathbf{W} \hat{\mathbf{y}}(\hat{\boldsymbol{a}}[n]). \end{aligned} \qquad 5.27$$

where $\mathbf{J}$ is the Jacobian matrix of the residual, of the current fit and latest data sample ($\hat{\mathbf{y}}[n](\hat{\boldsymbol{a}}[n]) - \mathbf{y}[n]$), with respect to the coefficients. When used to fit a function to the transfer function estimate, $\mathbf{y}[n] = \widehat{\mathbf{H}}[n]$ for complex fitting, or $\mathbf{y}[n] = \left|\widehat{\mathbf{H}}[n]\right|^2$ for Lorentzian fitting. In this case the curve fit to the spectrum should be weighted as $\mathbf{W} = diag(\mathbf{y}^{-1}[n])$, as the variance of each point is roughly proportional to $\left|\widehat{\mathbf{H}}\right|^2$, as discussed above. When fitting a function where all data points have the same variance, $\mathbf{W}$ reduces to the identity matrix.

The learning rate, $\mu$, ideally unity in the original Gauss-Newton algorithm, is used to stabilise the convergence of the algorithm in the presence of noise, *etc.* The concept of controlling $\mu$ based on the current convergence properties is the basis of many later additions to this algorithm, for example, the regularisation of Levenburg and Marquardt [37] can be seen to be a generalisation of the $\mu$ parameter. For the purposes of this simple realtime algorithm, where the change in

the spectral response with time can assumed to be relatively slow compared to the iteration rate, a scalar value of $\mu$ is perfectly sufficient.

Lorentzian fitting is appealing as it is still very accurate and high precision, as discussed above, yet does not require phase calibration. This eliminates the requirement for phase calibration in the multitone system. The development of this algorithm revealed a few tricks to using the Lorentzian function successfully. Firstly, rather than the 'book' definition, a normalised inverse quadratic works much better as a fit function:

$$\hat{\mathbf{y}}(\mathbf{a}) = \frac{a[1]}{1 + a[2]n + a[3]n^2}, \qquad n = 0 \ldots 1 = \frac{f - f_T}{\Delta f \cdot N}, \tag{5.28}$$

where $N$ is the number of tones, $f_T$ is the first tone frequency, and $\Delta f$ their frequency spacing (assumedly equal). This inverse quadratic function avoids a double-minimum problem due to having a coefficient squared (*i.e.,* $(a - n)^2$ would be a bad expression to have in a curve fit function as it has two solutions that give the same answer). Ensuring $n$ is always in the range [0,1] is also important for avoiding numerical divide-by-zero problems and multiple solutions due to the $n^2$. It also allows the same starting value to be used for all problems; $\mathbf{a}[0] = [1, -3, 3]^T$ almost always guarantees convergence. Furthermore, to eliminate any sensitivity to amplitude variations, the magnitude response estimate should be normalised before fitting *i.e.,*

$$\mathbf{y}[n] = \frac{\left|\hat{\mathbf{H}}[n]\right|}{\max\left|\hat{\mathbf{H}}[n]\right|}. \tag{5.29}$$

This ensures that the residual is always normalised to the same relative magnitude and is why a fixed $\mu$ is sufficient. This step has the added benefit that the absolute amplitude accuracy of the multitone hardware system is irrelevant. However, in order to invert for unloaded bandwidth, a 'calibration' value of $Q_e$ is needed as this cannot be estimated from the insertion loss in this case.

Conversion from these normalised fitted parameters to the eigenvalue estimates (in Hz) is then simple:

$$\widehat{\sigma_1} = \Delta f \cdot N \cdot \frac{\sqrt{|4|\hat{a}[3]| - (\hat{a}[2])^2|}}{|\hat{a}[3]|}, \qquad \hat{f}_1 = f_T - \Delta f \cdot N \frac{\hat{a}[2]}{2\hat{a}[3]} \tag{5.30}$$

A version of this fast Lorentzian curve fit was implemented in LabVIEW. The Jacobian **J** was expanded into is analytical form, allowing for its efficient calculation,

$$\left[\frac{\hat{\mathbf{y}}(\mathbf{a}) - \boldsymbol{y}}{1 + a[2]\mathbf{n} + a[3]\mathbf{n}^2} \quad \frac{-(\hat{\mathbf{y}}(\mathbf{a}) - \boldsymbol{y})a[1]\mathbf{n}}{(1 + a[2]\mathbf{n} + a[3]\mathbf{n}^2)^2} \quad \frac{-(\hat{\mathbf{y}}(\mathbf{a}) - \boldsymbol{y})a[1]\mathbf{n}^2}{(1 + a[2]\mathbf{n} + a[3]\mathbf{n}^2)^2}\right]^T \tag{5.31}$$

The resulting was reduced to $28N + 40$ multiplications and $13N + 20$ addition operations per iteration, or $19N + 40$ multiplications without using weighting.

The performance of the algorithm was evaluated using some simple numerical simulation experiments. The first experiments were static simulations, based on initialising the algorithm to the same starting coefficients each time and performing 100 iterations at $\mu = 0.1$, then evaluating the estimation error $(\widehat{\sigma_1} - \sigma_1)^2$ and $\left(\hat{f} - f_1\right)^2$. A 31-tone excitation was assumed. The first of these experiments, Figure 5.29, was for no input noise – the error displayed by this simulation is the algorithm convergence error after the finite time of 100 iterations. Close to the starting point, the error is very low and comparable to the numerical precision. However, the farther the coefficients are from the starting value the higher the error. This is reflected in the red regions of the plots. It is equivalent to the dynamic convergence seen in the adaptive filter in response to a step change – the bigger the step change, the more iterations are needed to reach a given error level. In practice, if the rate of change of eigenvalue is limited, this source of error would be very small.

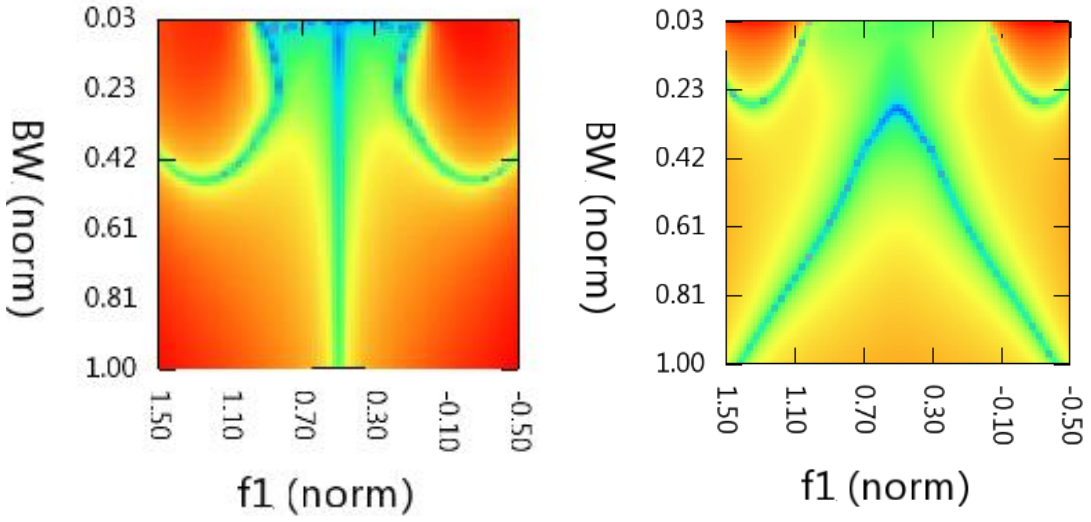



*Figure 5.29: Performance of the fast Lorentzian fit algorithm with no noise for various bandwidth and resonant frequency values. Left shows the error in the resonant frequency estimate, from blue at -150 dB through to red at -60 dB. Right shows the error in bandwidth estimate, from blue at -150 dB through to red at -40 dB.*

Figure 5.30 is a Monte Carlo simulation of performance at an input SNR of 80 dB at the mid-bandwidth value of 0.5 (the SNR is assumed to be inversely proportional to the bandwidth). This time the algorithm was run for 1000 iterations. This therefore represents the steady-state performance of the algorithm independent of convergence time. This shows that there is a modest SNR gain of about 20 dB, in general, roughly in agreement with the earlier simulations. Evidently, when the resonance peak is outside of the measurement band $f_{1,n} > 1$ or $f_{1,n} < 0$, a significant degradation in accuracy is seen.

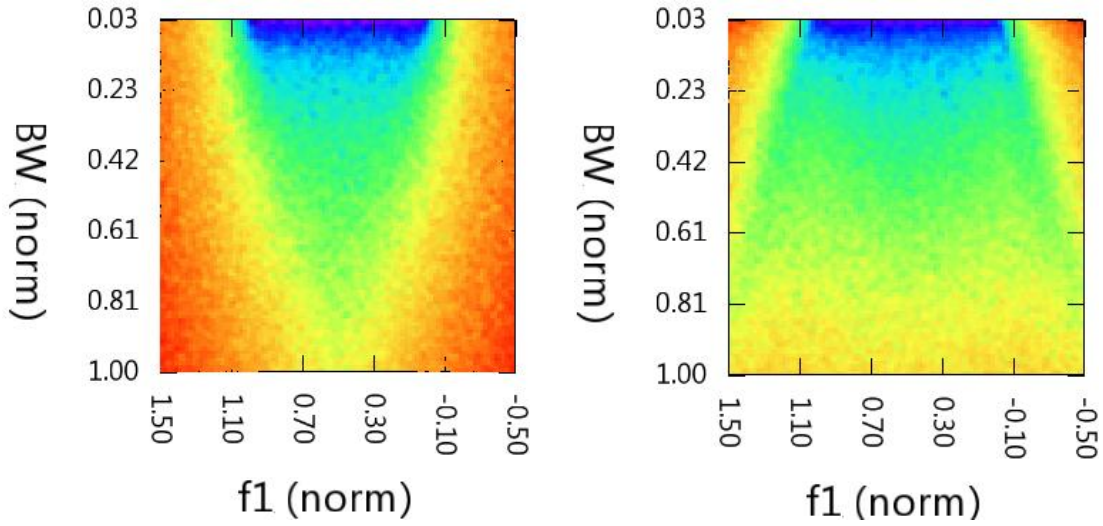



*Figure 5.30: Performance of the fast Lorentzian fit algorithm with an input SNR of 80 dB (100 averages). Left shows the error in the resonant frequency estimate, from blue at -150 dB through to red at -60 dB. Right shows the error in bandwidth estimate, from blue at -150 dB through to red at -40 dB.*

### 5.3.2 Nonparametric, neural network approaches

The curve fitting method is very powerful, but it even with an optimised implementation, it is still quite computationally intensive. Furthermore, due to the limited convergence rate, it is unsuitable for use on the most rapidly varying sensors. Whilst it is possible to repeat iterations with the same data to improve convergence rate, this also increased the computational burden by the same level – not an efficient trade-off.

Therefore, what is needed is a direct inversion of the eigenvalue function – in other words an algorithm that, in one-step, evaluates the function $\hat{\mathbf{p}}[n] = g(\hat{\mathbf{H}}[n])$. As discussed in Section 3.6.1, artificial neural networks (ANN) are powerful methods to find complex multi-dimensional functions based on input training data. A multilayer neural network could approximate the eigenvalue response inverse function in a single time sample:

$$\hat{\mathbf{p}}[n] = \mathbf{A}^T \boldsymbol{f}\left(\mathbf{B}^T \boldsymbol{f}\left(\mathbf{C}^T \hat{\mathbf{H}}[n]\right)\right) \approx g\left(\hat{\mathbf{H}}[n]\right). \qquad 5.32$$

The weight matrices **A**, **B** and **C** are taught to the ANN from numerical simulations of sensor response over a large, fully-sampled grid over the anticipated range of eigenvalue parameters. The functions $\boldsymbol{f}(.)$ are the neural network activation functions and may be of the 'sigmoid' type; $y = 2/(1 + e^{-2x}) - 1$, or some more computationally-efficient function such as a Look Up Table (LUT) or a truncated power series representation. Assuming a LUT implementation with zero multiplications, the total computational complexity for a two-layer neural network in feed-forward mode like this is $N \cdot N_{H1} \cdot N_{H2} \cdot 2$ multiplications. Depending on the number of hidden units $N_{Hi}$ needed to learn the function with sufficient accuracy, this may be more efficient than a conventional curve-fitting approach.

One of the strengths of ANN learning is that a network can be trained with realistic data; *e.g.,* sensor input and output noise, and the normalisation of the transfer function amplitude. It can also be trained to be insensitive to non-ideal effects such as connecting transmission line length and coupling reactance.

It has been demonstrated that adaptive filters are an efficient method for estimating the sensor impulse response under rapidly varying conditions. Thus, paired together, the adaptive filter and neural network could be an enabling combination for the highest speed MRS measurements. In this configuration, the ANN follows the adaption algorithm, using as its input the time-domain filter weights that have been filtered to reject non-excited frequencies (this function can be incorporated into the multiplication of the first ANN weight matrix, **C**, and thus does not constitute any additional computational complexity). The weights can also optionally be low-pass filtered and decimated if necessary before the ANN inversion.

Furthermore, the eigenvalue response function can also be incorporated into the ANN function. Hence this system can output not only the eigenvalue $\hat{\mathbf{p}}[n]$, but additionally (or alternatively) the desired measurand parameter estimate $\hat{\boldsymbol{\theta}}[n]$. This last option is application specific. However, being merely software, new ANN

coefficients can easily be loaded to perform sensing with different resonators and in different application contexts without significant change to the architecture of the readout system software.

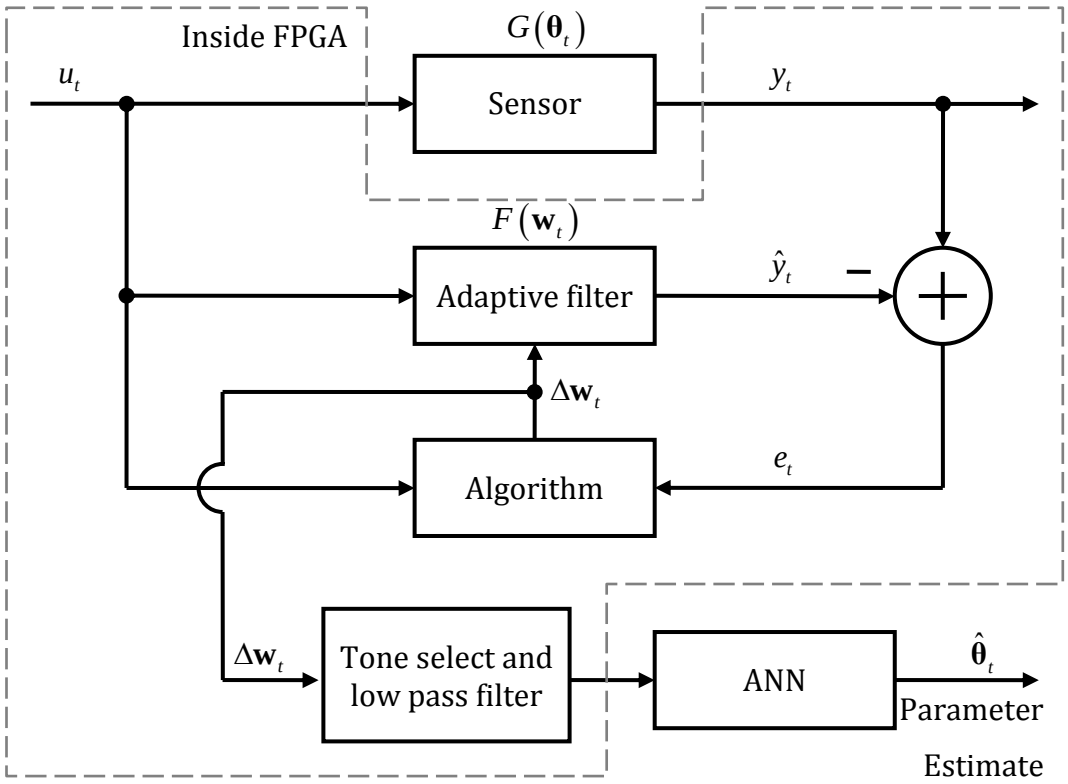



*Figure 5.31: Complete real time parameter estimator made from an adaptive filter followed by an artificial neural network*

To demonstrate this system through simulation, a neural network was trained with data simulated from an adaptive filter given plenty of time to converge (*i.e.,* as if the sample properties were static). The operating point was assumed when the resonator is centred on the middle of the multitone signal. The neural network was trained to recognise both increases and decreases in resonant frequency and bandwidth over a regularly sampled grid. Half the training dataset was noiseless and half had additive noise at a level of -75dB (this was uniform across the change in bandwidth) so the network learnt robust noise rejection.

The neural network was trained with 2000 samples, generated by an optimised Latin square sample of the bandwidth-frequency space [38]. This ensures that the sample distribution is much more evenly spread than if drawn from a simple 2D uniform distribution, but is still essentially nondeterministic. This is good for training, as multiple samples from this same distribution can be used to compare the generalisation properties of the neural network (*i.e.,* to unseen data) in a statistically rigorous way [10].

The number of adaptive filter weights simulated was 32; these were divided into real and imaginary (the neural network algorithm in the Matlab toolbox does not work with complex data), giving a total of 64 inputs. A three-layer neural network architecture was chosen, having two hidden layers each with hyperbolic tangent nonlinearities. Ten hidden neurons were chosen for each hidden layer. This number was reached by guesswork, although trials over all possible combinations could be done given sufficient time. Bayesian regularisation was chosen as the training algorithm [39], this algorithm, whilst computationally intensive and very slow, is well suited to this sort of regression problem and gives excellent generalisation performance.

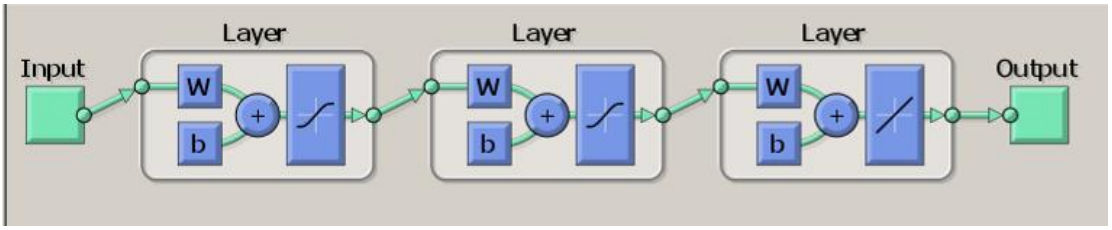



*Figure 2.9: Neural network toolbox representation of the network*

The error shown below is the absolute relative validation error, $e = |(\hat{\mathbf{p}} - \mathbf{p})/\mathbf{p}|$. This error is based on the evaluation of ANN performance with new samples not used to train the neural network. As such, it is a good measure of the generality of the network; it is also a fair approximation to $1/SNR$. The samples are shown by small grey dots. The value predicted by the neural network is actually plotted over the true value, but the error is so small that no difference can be discerned by eye. As seen by the contour plot, the error surface has quite a large amount of ripple. This may imply that too many hidden units were used. However, the average error

is around −80dB, which corresponds to an uncertainty of ± 0.01% (100 Hz in 1 MHz) in the estimate of bandwidth or resonant frequency change. From the plot, it is apparent that the SNR degrades under smaller changes. This is because the change is small in these areas and the performance measure is relative. This could be improved by retraining with many more samples from these ranges.

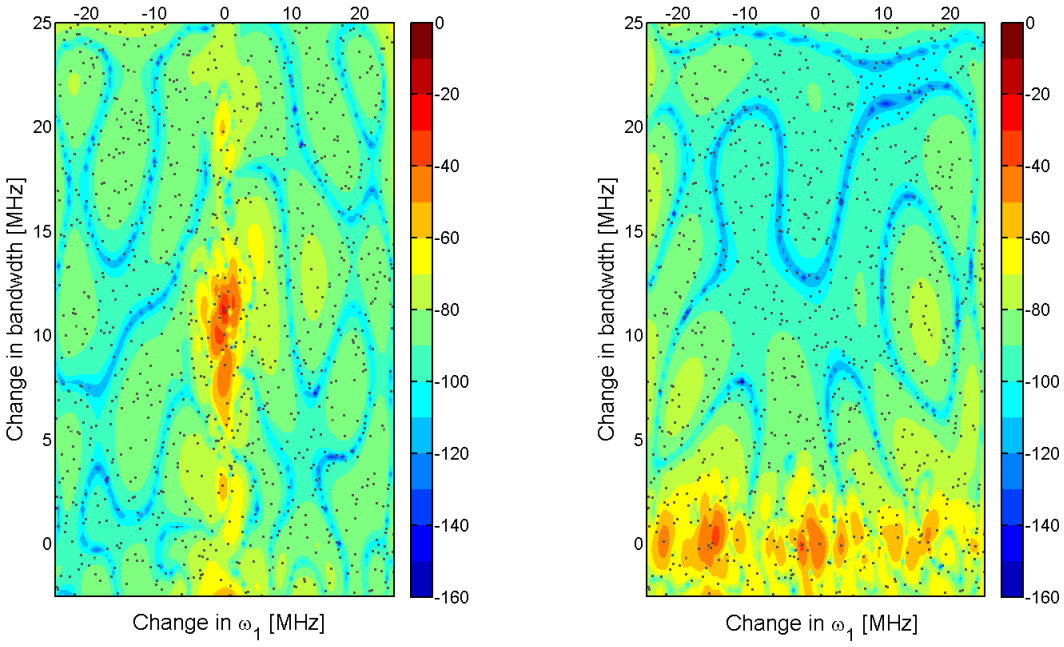



*Figure 5.32: Validation error after neural network prediction of resonant frequency and bandwidth change. On the left is the absolute relative error in the estimate of the change in resonant frequency, scale is in dB (-40dB is an error of ±1%). On the right is the absolute relative error in the estimate of the change in bandwidth, same scale*

In summary, the performance of the ANN method is comparable to that of the fast Lorentzian method, therefore providing a neat and flexible solution to the problem of eigenvalue estimation. However, this combined algorithm remains to be tested practically, and a comprehensive analysis of the error and statistical performance of the combined system remains an outstanding task.

## 5.4 Chapter references


[1] L. Cohen, “Generalization of the Wiener-Khinchin theorem,” *IEEE Signal Processing Letters*, vol. 5, no. 11, pp. 292-294, Nov. 1998.

[2] K. Godfrey, *Perturbation signals for system identification*. Prentice Hall, 1993.

[3] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*. Prentice Hall, 2010.

[4] R. G. Lyons, *Understanding Digital Signal Processing*. Prentice Hall, 2010.

[5] E. Evans, D. Rees, and L. Jones, “Nonlinear disturbance errors in system identification usingmultisine test signals,” *IEEE Transactions on Instrumentation and Measurement*, vol. 43, no. 2, pp. 238-244, Apr. 1994.

[6] M. Schroeder, “Synthesis of low-peak-factor signals and binary sequences with low autocorrelation (Corresp.),” *IEEE Transactions on Information Theory*, vol. 16, no. 1, pp. 85- 89, Jan. 1970.

[7] E. Van der Ouderaa, J. Schoukens, and J. Renneboog, “Peak factor minimization using a time-frequency domain swappingalgorithm,” *IEEE Transactions on Instrumentation and Measurement*, vol. 37, no. 1, pp. 145-147, Mar. 1988.

[8] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *, IEEE International Conference on Neural Networks, 1995. Proceedings*, 1995, vol. 4, pp. 1942-1948 vol.4.

[9] M. E. H. Pedersen and A. J. Chipperfield, “Simplifying Particle Swarm Optimization,” *Applied Soft Computing*, vol. 10, no. 2, pp. 618-628, Mar. 2010.

[10] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley, 2001.

[11] B. Widrow and M. Hoff, “Adapting switching circuits,” in *1960 {{\sc ire wescon}} Convention Record*, 1960, pp. 96-104.

[12] H. Robbins, “A Stochastic Approximation Method,” *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400-407, Sep. 1951.

[13] S. S. Haykin, *Adaptive filter theory*. Prentice Hall, 2002.

[14] A. H. Sayed, *Fundamentals of adaptive filtering*. Wiley-IEEE, 2003.

[15] J. G. Proakis, *Digital communications. 1995*. McGraw-Hill.

[16] Y. Bar-Shalom and U. of C., Los Angeles. University Extension, *Multitarget-multisensor tracking: applications and advances*. Artech House, 1992.

[17] J. Naylon and A. Porch, “UK patent application number GB1109685.6: Apparatus and method for estimating a characteristic of a microwave resonant device,” 2011.

[18] F. Gustafsson, *Adaptive filtering and change detection*. Wiley, 2000.

[19] F. H. Raab et al., “Power amplifiers and transmitters for RF and microwave,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 50, no. 3, pp. 814-826, Mar. 2002.

[20] M. O’Droma et al., “Developments in predistortion and feedforward adaptive power amplifier linearisers,” in *Gallium Arsenide and Other Semiconductor Application Symposium, 2005. EGAAS 2005. European*, 2005, pp. 337-340.

[21] A. Lohtia, P. Goud, and C. Englefield, "An adaptive digital technique for compensating for analogquadrature modulator/demodulator impairments," in , *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, 1993*, 1993, vol. 2, pp. 447-450 vol.2.

[22] E. Acar and S. Ozev, "Digital calibration of RF transceivers for I-Q imbalances and nonlinearity," in *25th International Conference on Computer Design, 2007. ICCD 2007*, 2007, pp. 512-517.

[23] F. Harris, "Digital filter equalization of analog gain and phase mismatch inI-Q receivers," in *, 1996 5th IEEE International Conference on Universal Personal Communications, 1996. Record*, 1996, vol. 2, pp. 793-796 vol.2.

[24] D. Kajfez and E. J. Hwan, "Q-Factor Measurement with Network Analyzer," *IEEE Transactions on Microwave Theory and Techniques*, vol. 32, no. 7, pp. 666- 670, Jul. 1984.

[25] E. L. Ginzton, *Microwave measurements*. McGraw-Hill, 1957.

[26] En-Yuan Sun and Shuh-Han Chao, "Unloaded Q measurement-the critical points method," *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, no. 8, pp. 1983-1986, Aug. 1995.

[27] S. Shahid, J. A. . Ball, C. G. Wells, and P. Wen, "Reflection type q-factor measurement using standard least squares methods," *IET Microwaves, Antennas & Propagation*, vol. 5, no. 4, pp. 426-432, Mar. 2011.

[28] D. Kajfez, *Q Factor Measurements Using MATLAB*. Gardners Books, 2011.

[29] D. Kajfež, *Q factor*. Vector Fields, 1994.

[30] D. Kajfez, "Random and systematic uncertainties of reflection-type Q-factor measurement with network analyzer," *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, no. 2, pp. 512-519, 2003.

[31] D. Kajfez, S. Chebolu, M. R. Abdul-Gaffoor, and A. A. Kishk, "Uncertainty analysis of the transmission-type measurement of Q-factor," *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 3, pp. 367-371, 1999.

[32] D. Kajfez, *Q Factor Measurements in MATLABŽ: Software and User's Guide*. Artech House, 2011.

[33] P. J. Petersan and S. M. Anlage, "Measurement of resonant frequency and quality factor of microwave resonators: Comparison of methods," *Journal of Applied Physics*, vol. 84, no. 6, pp. 3392-3402, 1998.

[34] R. Inoue, K. Miwa, H. Kitano, A. Maeda, Y. Odate, and E. Tanabe, "Highly accurate and real-time determination of resonant characteristics: complex linear regression of the transmission coefficient," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 9, pp. 2163- 2168, Sep. 2004.

[35] J. Gao, "The physics of superconducting microwave resonators," 2008. [Online]. Available: http://thesis.library.caltech.edu/2530/. [Accessed: 04-Oct-2011].

[36] A. Zielesny, *From Curve Fitting to Machine Learning: An Illustrative Guide to Scientific Data Analysis and Computational Intelligence*. Springer, 2011.

[37] D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431-441, Jun. 1963.

[38] P. L. Munholland and J. J. Borkowski, “Simple Latin Square Sampling + 1: A Spatial Design Using Quadrats,” *Biometrics*, vol. 52, no. 1, pp. 125-136, Mar. 1996.

[39] D. J. C. MacKay, “A Practical Bayesian Framework for Backpropagation Networks,” *Neural Computation*, vol. 4, no. 3, pp. 448-472, Oct. 2011.

# 6 Verification

To demonstrate and validate the measurement of dynamic perturbation signals using a microwave resonant sensor (MRS), a range of experiments were performed. Two of these aimed to verify the accuracy of the multitone readout system pushed at high sampling rates; the capture of a falling liquid drop, in Section 6.1, and the measurement of a gas-segmented flow stream in Section 6.2. Neither of these experiments seems to have been demonstrated before, and both represent challenging measurement environments that would be difficult or impossible for traditional MRS readout systems to handle with good precision and accuracy simultaneously.

A project to monitor and control microwave heating effects in miniature flow systems provides a good demonstration of the use of temporal measurement, in Section 6.3. In Section 6.3, the use of temporal MRS measurement for dynamic capillary filling experiments is investigated. These experiments are all proof-of principal demonstrations; however, each shows potential as a sensor system in its own right that, with further development, might find application in metrology, chemical synthesis, point-of-care diagnostics, or remote sensing, for example.

For the purposes of verification, the experiments were devised so that a reasonably good prediction of the results could be made by an independent method, or by sound theoretical analysis. An agreement with prediction therefore serves to verify temporal MRS measurements and the multitone readout system. All these experiments, therefore, took a known 'reference material' and used it in one-way or another to create a dynamic signal. Liquids are ideal for this task – simple, pure solvents are well-characterised for permittivity against frequency and temperature by groups such as the National Physical Laboratory, UK, giving a traceable standard by which to compare measurements. Thus, liquids and liquid metrology is a strong theme throughout this work and in potential applications of these sensor systems.

## 6.1 Capturing a falling dielectric drop

An existing aluminium cylindrical $TM_{010}$ cavity resonator designed for the measurement of tubes containing liquid or powdered materials [1] was used as a sensor for a dynamic measurement verification experiment where an invariable dielectric liquid was dripped at a steady rate through the resonator body. The cavity was designed with a 5 mm circular aperture centred in its top and bottom surfaces. A microcapillary-based dropper was aligned precisely with these holes so that each droplet fell along the axis of the resonator, with the time between entering and exiting the resonator being about 40 ms. This created a repeatable, high speed dynamic perturbation signal having a peak resonant frequency slew-rate of about 30 MHz/s with which to test the prototype system.

The reference liquid, for which both ethanol and methanol were used (HPLC grade, Sigma Aldrich, UK), was delivered to the dropper at constant rate by a syringe pump (KDS200, KD Scientific) which maintained an accurate flow rate of 100 µl/min. The dropper was made from a length of accurately cleaved, clean polyether ether ketone (PEEK) capillary with an inner diameter (ID) of 150 µm. Due to its small dimensions, the flow regime in the capillary was laminar and the drop formation highly regular. The process of drop formation is highly controlled: gradually the droplet increases in volume under the slow, steady and pulse-free flow conditions provided by the pump, with surface tension and electrostatic attraction to the hydrophobic capillary being the only competing forces to gravity. Once the drop reaches a certain size, it is no longer in contact with the capillary and is held only by cohesive forces to the remaining liquid. Eventually the weight of the forming droplet overcomes surface tension and the droplet falls. When carefully controlled, this process is highly predictable, and is a long established way of measuring surface tension using a specialised pipette known as a stalagmometer [2],[3] . This technique relates the surface tension of a liquid to a droplet's mass *m* and the radius of the capillary orifice *r* [4] :

$$\gamma_e = \frac{mg}{2\pi r f_c}, \tag{6.1}$$

where $g = 9.81\ ms^{-2}$ is the gravitational constant and $f_c$ is a correction factor of order 1 that accounts for imperfections in a specific aperture's geometry – generally this factor is eliminated in practice by comparing to a reference liquid such as water. Although this capillary is not flat-bottomed as a stalagmometer, a similar relationship to the above should hold in this case. Furthermore, given the small size of the drop and the low viscosity of the alcohols, it is expected that the drops should form a neck only for a very short time and without creating any satellite droplets following the main drop. Thus by the time the drop enters the cavity it is expected to be isolated and spherical. Whether these assumptions are true in practice could be verified with a high-speed video camera setup if required, such as that demonstrated in [5].

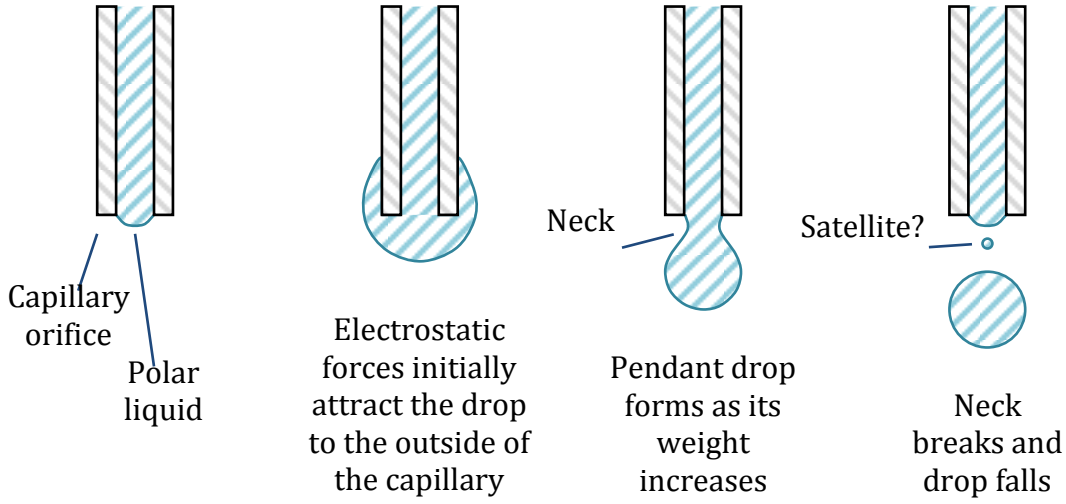



*Figure 6.1: Formation of a liquid drop at the tip of a hydrophobic microcapillary*

Once detached the droplet falls as a particle under free fall, described by Newtonian mechanics and therefore having a velocity independent of mass. This is easily verified by calculating the terminal velocity for a spherical object. Considering that the particle is large enough to be in a high Reynolds number regime, this is given by [6]

$$v_t = \sqrt{2mg/\rho C_D A}, \qquad 6.2$$

where $m$ is the droplet mass, $\rho \approx 1.2\ \mathrm{kgm^{-3}}$ is the density of air, $C_D \approx 0.47$ (dimensionless) is the drag coefficient of a sphere and $A = \pi r^2$, $r \approx 880$ μm is the cross-sectional normal area. Thus for these droplets the terminal velocity is over 5.7 $\mathrm{ms^{-1}}$, well in excess of the 1.1 $\mathrm{ms^{-1}}$ that the droplets are expected to reach by the time they exit the resonator. Thus using the ballistic approximation, which predicts the time taken to fall a distance $h$ is $\sqrt{2h/g}$, the transit time of the drops should be constant at around $\Delta t = 44.9$ ms.

It is therefore reasonable to assume that the main experiment parameters; liquid composition, droplet size, interval and freefall time are all well controlled, allowing the performance of the system to be reliably assessed with a predictable and repeatable time-varying perturbation. Figure 6.2 showed the experimental setup. A short length of PEEK capillary of inner diameter (ID) 150 μm and outer diameter (OD) 360 μm (Upchurch Scientific, UK) was threaded through a sleeve of OD 1/16". This created a tight seal in the conical Fingertight fitting (yellow-grey) and union (black) that interfaced the dropper assembly to the main tubing. Standard 500 μm ID, 1/16" OD Teflon tubing then connected this assembly to a gas tight glass 10 ml syringe for precision dispensing.

The syringe pump employed here works by rotating a lead-screw driven pusher using a stepper motor assembly, providing an accurate flow as long as the diameter of the syringe is known precisely (accuracy better than 1 %, repeatability better than 0.1 %) [7]. The syringe pump was calibrated by measuring the syringe plunger diameter with a micrometer (14.57 mm in this case). This information is input into the pump, ensuring accurate conversion from pusher velocity to volumetric flow rate. The flow rate was set to 100 μl/min, allowing the drop to still form freely and naturally whilst ensuring that the drop interval was not excessive.

The dropper assembly was suspended vertically above the cylindrical resonator and aligned to its main axis (see Figure 6.10). Inserted through the resonator body was a 5 mm diameter thin-walled plastic straw. This protected the inside of the

resonator from contamination in the event of misaligned drop. Being of constant cross section, minimal volume, and low permittivity it had minimal effect on the resonance. The resonator was connected by short cables to the second implementation of the PXI-based multitone prototype system as discussed in the Hardware chapter. Data were captured, processed, and streamed to file in realtime by the same LabVIEW program used to implement the processing algorithms.

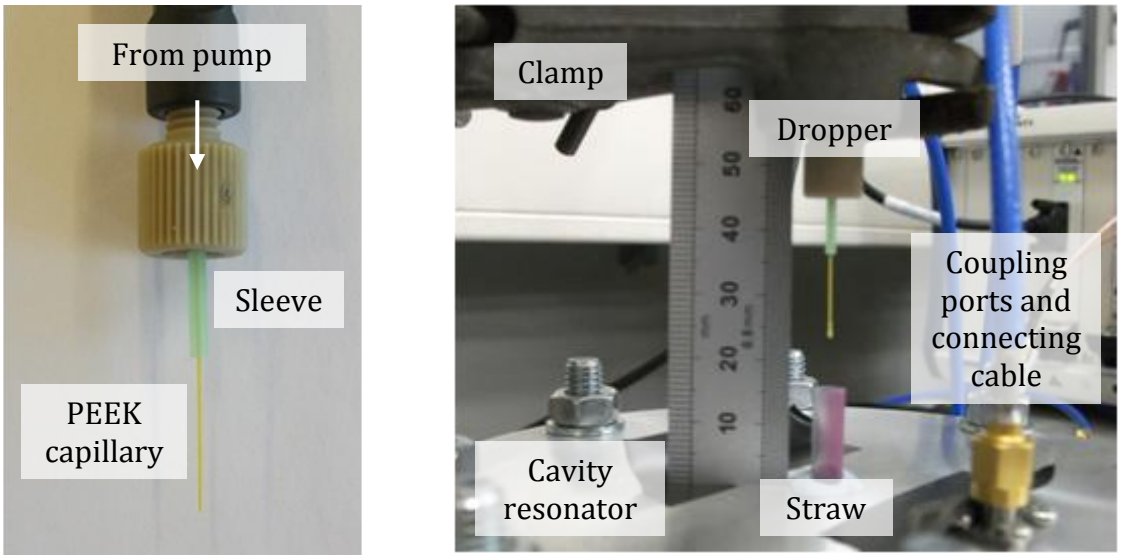



*Figure 6.2: Photograph of the dropper assembly (left) and the experimental setup (right)*

### 6.1.1 Resonant sensor theory and characterisation

In order to extract useful sample measurements from this experimental setup, the microwave resonant sensor, *i.e.,* the $TM_{010}$ cavity, needs to be characterised. As discussed in the Theory chapter, this will allow measurements of resonant frequency and bandwidth to be converted to complex permittivity through the inversion of the resonator eigenvalue response function.

The general perturbation Equation 3.14 describes the link analytically between complex resonator eigenvalue $p$ and a desired measurand, $\theta$, which in this case is the complex permittivity of the droplet:

$$\frac{p(\theta)-p_0}{p_0}=\frac{-\sum_{i=1}^{n} e_i\Delta\varepsilon_i(\theta)-h_i\Delta\mu_i(\theta)-\sum_{k=1}^{m} h_{S,k}\Delta\eta_k(\theta)-\sum_{j=1}^{p} e_j\Delta\tilde{\varepsilon}_j(\theta)}{\sum_{i=1}^{n} e_i\varepsilon_i(\theta)+h_i\mu_i(\theta)+\sum_{k=1}^{m} h_{S,i}\eta_k(\theta)+\sum_{j=1}^{p} e_j\tilde{\varepsilon}_j(\theta)}. \quad 6.3$$

The sample in this case is assumed small and spherical, and will depolarise in the resonator electric field because the axial electric field lines in the $TM_{010}$ mode will cross the dielectric boundary of the sample. Therefore, for the single sample region, the best energy term to use out of those in the above general perturbation equation is the dipole moment term, *i.e.,* the fourth summation term, $e_j\Delta\tilde{\varepsilon}_j(\theta)$, on the numerator. Taking the reference to be air so that $\Delta\tilde{\varepsilon}_j(\theta)\approx\tilde{\varepsilon}_j$, this gives a simplified energy term for a dipole in a uniform applied field:

$$e_j\tilde{\varepsilon}_j=E(0)\cdot p^*(\theta)=V_j\,E(0)\cdot\frac{(\varepsilon_i-\varepsilon_e)}{\varepsilon_e+N(\varepsilon_i-\varepsilon_e)}E^*(\theta). \quad 6.4$$

There is only one non-sample region, the cavity body, and only one sample region. Therefore, all summations reduce to single values. In the cavity air region $e_i\triangleq\int E(0)\cdot E^*(\theta)\;dV_i$ and $h_i\triangleq\int H(0)\cdot H^*(\theta)\;dV_i$, but it can be assumed that $E(0)=E(\theta)$ because the sample is so small it should cause an insignificant change to the field distribution. Because of this assumption, the equipartition theorem can also be used to state that $e_i=h_i$ . For the $TM_{010}$mode, the only non-zero field components are the azimuthal magnetic field, which varies only with radius, and the vertical electric field, which also varies only with radius. This simplifies the field integral calculations for these terms greatly. Thus,

$$e_i=h_i=\int|E(0)|^2\;dV=2\pi\ell\,E_0^2\int_0^a r\,J_0^2(\alpha r/a)\;dr=V_c\,E_0^2\,J_1^2(\alpha), \quad 6.5$$

where $V_c$ is the cavity volume [8],[9], Here the constant $\alpha=2.404\,825\,557\,695\ldots$ is the solution to $J_0(\alpha)=0$; the first positive real root of the Bessel function of the first kind of zeroth order. The software package Maple® (Maplesoft Inc., ON, Canada) can be used to find any such root if other cylindrical resonator modes are used instead: in this case the constant $J_1^2(\alpha)\cong 0.269\,514\,124$. Furthermore, for an air filled resonator it can be assumed that the relative $\varepsilon_i(0)\cong 1$. Even though the permittivity of air may be slightly higher than this, depending largely on humidity,

the error introduced by this assumption is significantly less than cavity dimension uncertainties, *etc.*

$$\frac{p(\theta) - p_0}{p_0} = \frac{-\, e_j \Delta\tilde{\varepsilon}_j(\theta)}{2e_i + \, e_j \Delta\tilde{\varepsilon}_j(\theta)}. \tag{6.6}$$

Assuming the drop is centred exactly at the electric field maximum so that $E(0) = E_0$, and assuming the drop is perfectly spherical so that $N = 1/3$, gives the simplified energy factor for the depolarised dielectric droplet:

$$e_j \Delta\tilde{\varepsilon}_j = 3V_d E_0^2 \cdot \frac{\varepsilon_d - 1}{\varepsilon_d + 2}, \tag{6.7}$$

where $V_d$ is the drop volume and $\varepsilon_d$ its complex relative permittivity. Thus the fully simplified linear resonator perturbation equation for the droplet in a $\mathrm{TM}_{010}$ mode resonator is

$$\frac{p(\theta) - p_0}{p_0} = \frac{-3V_d(\varepsilon_d - 1)}{2V_e(\varepsilon_d + 2) + 3V_d(\varepsilon_d - 1)}, \tag{6.8}$$

where $V_e = J_1^2(\alpha)V_c$ is the effective cavity volume. This function is plotted in Figure 6.3 below:

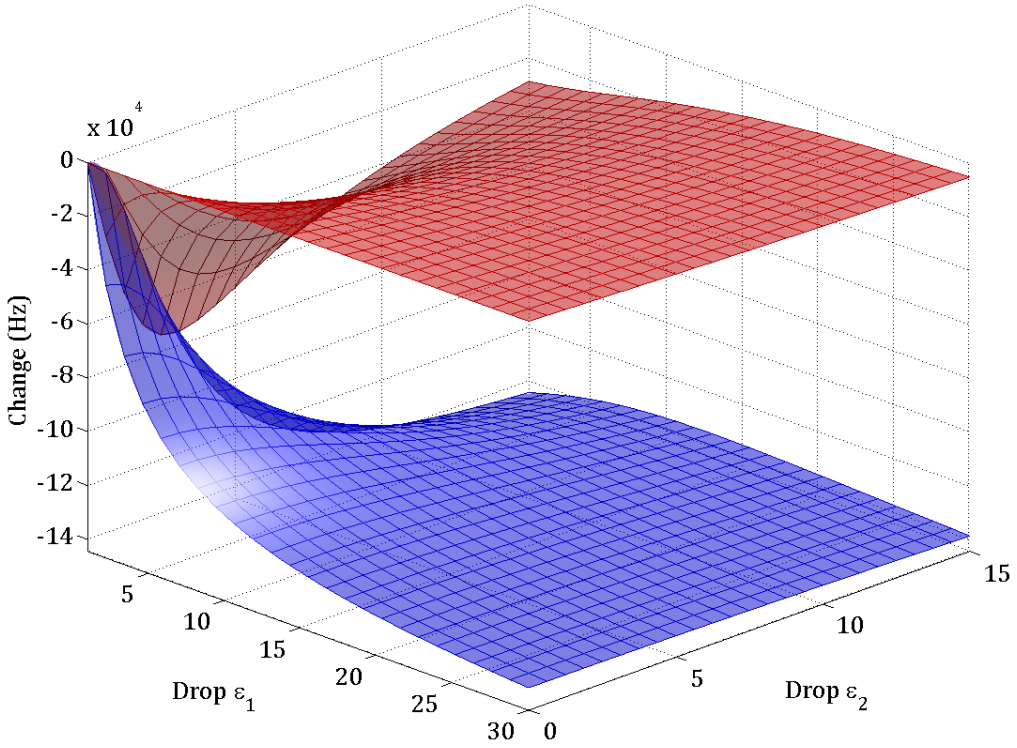



*Figure 6.3: Predicted change in $\omega_0$ (blue) and $\sigma_0$ (red) against drop permittivity according to Equation 6.8 for the $TM_{010}$ mode cavity. A drop volume of 2.84 μl was assumed. Note that, by the definition used here, the '3dB bandwidth' $B_0 = -2\sigma_0$, the change in which is positive – as expected for lossy sample.*

This function is clearly nonlinear due to the depolarising effect, and this sensor is not an optimal way of measuring liquid permittivity because of it. The predicted change above is only valid when the droplet is in a central position, at the mid-point of its fall through the resonator. However, the assumption of uniform electric field along the cavity axis is clearly invalidated close to the holes which allow the drop to enter and leave the cavity. In addition, the effect of drop depolarisation fields may cause nonlinear perturbation effects when the drop is close to the cavity ends.

According to [10] the permittivity of methanol at 25°C and 2.5 GHz is $22.15 \pm 0.08 - j13.20 \pm 0.06$ (intervals given at 95% confidence). The predicted reduction in resonant frequency for methanol is therefore $136 \pm 1.3$ KHz and unloaded bandwidth increase of $7.88 \pm 0.14$ KHz, based on an uncertainty of $\pm 0.2$ mm in

cavity dimensions (estimated from measurement at various different positions – this relatively large variation may have been caused by warping due to the resonator's assembly method). For ethanol, $7.08 \pm 0.06 - j6.83 \pm 0.03$ at the same temperature and frequency, this perturbation estimate is $119 \pm 1.3$ KHz for resonant frequency and $24 \pm 0.46$ KHz for bandwidth. These estimates will provide a reference by which to verify the accuracy of the dynamic measurement.

The inverse of this expression gives the estimated complex $\varepsilon_d$ as a function of the measured eigenvalue:

$$\hat{\varepsilon}_d = \frac{p(\theta)(3V_d - 4V_e) + 4\,p_0 V_e}{p(\theta)(3V_d + 2V_e) - 2\,p_0 V_e}, \qquad 6.9$$

In this expression it is necessary to correct for the effects of loading using the approximation given in the Theory chapter, $\sigma_0 \cong \sigma_1 - \omega_1/2Q_e$, when using measured results. This is particularly important here as the resonator is strongly coupled; ignorance of this would therefore lead to a significant error.

In order to determine the external $Q$, a 2-port, fully calibrated VNA measurement of the empty resonator was made. An 8-term spectral model was fitted to these data based on a first order expansion of the effect of the electrical length of the capacitive coupling probes beyond the calibrated reference plane, which was about 12.2 mm for each probe:

$$\hat{y} = [1 + f'(a_1 + ja_2)]\left(a_3 + ja_4 + \frac{\left(\frac{1}{2} + a_5 + a_6\right)a_8}{a_7 + j(f' - a_8)}\right), \qquad f' = \frac{f - f_{centre}}{f_{span}} + \frac{1}{2}, \qquad 6.10$$

$f'$ are the normalised frequency points of the sweep (normalisation improves numerical stability and convergence). This curve fit (and all subsequently described curve fitting) was performed using the Matlab® Optimization Toolbox™ (The MathWorks Inc., MA, USA). The toolbox function `lsqnonlin` was used. This implements a Levenburg-Marquardt algorithm, as discussed in the Software chapter, to minimise the 2D objective $C = [\Re(\hat{\mathbf{y}}(\mathbf{a}) - \mathbf{y}),\ \Im(\hat{\mathbf{y}}(\mathbf{a}) - \mathbf{y})]^T$, where $\mathbf{y}$ is a vector of measured complex scattering parameters, $\hat{\mathbf{y}}$ is the fitted function

estimate and **a** the parameter vector. This circumvents the limitation of the fit algorithm to real-domain functions, allowing full use of the complex scattering parameter data whilst avoiding a nonlinear transformation to power, phase, *etc*. A nonlinear transformation would invalidate the assumption of normally distributed noise inherent in least squares curve fitting. Because of the way the function is defined, an initial coefficient vector $\mathbf{a}_0 = [0, 0, 0, 0, 0, 0, 0.05, 0.5]^T$ always provides a good starting point for the fit; the first six coefficients being zero in the case of zero imperfections (*i.e.,* free from coupling reactance, connecting line length, *etc*.).

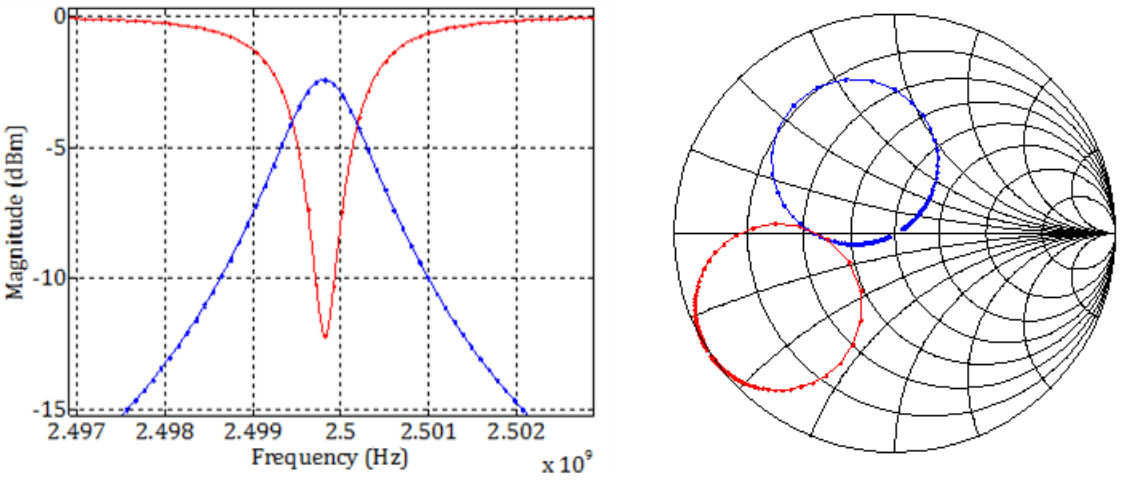



*Figure 6.4: Calibrated reference measurement of the $TM_{010}$ resonator showing $S_{21}$ (blue) and $S_{11}$ (red) in magnitude (left) and Smith chart (right) representations; the data are shown by points and the curve fit by a continuous line.*

The mean squared error of this fitted function, $e_{mse} = \nu^{-1}\|\mathbf{y} - \hat{\mathbf{y}}\|^2$, where ν is the number degrees of freedom of the fit. This figure of merit is extremely small for a single sweep due to the very low trace noise of the VNA – about $10^{-8}$. This value can be transformed into an estimate for the upper and lower bounds on the derived coefficients using the following relationship:

$$\hat{\mathbf{a}} = \mathbf{a} \pm t \cdot \sqrt{diag\{e_{mse} \cdot (\mathbf{J}^T\mathbf{J})^{-1}\}}, \qquad 6.11$$

where $t$ is the inverse of Student's cumulative distribution function for $\nu = 2N - 8$ degrees of freedom, $N$ being the number of samples (the factor of 2 comes from the

real and imaginary components which are assumed to be independent). **J** is the Jacobian matrix of the fitted **y** values with respect to the fitted coefficient vector – this is something most curve fitting algorithms can optionally return numerically; although for this function it does also have an analytical representation. The normalised coefficient vector **a** is related to useful resonator parameters by the following series of relationships:

$$\begin{aligned} f_1 &= f_{centre} + f_{span}\left(a_8 - \tfrac{1}{2}\right), & K_1 &= \frac{a_8}{a_7}\left(\tfrac{1}{2} + a_5 + a_6\right), \\ \sigma_1 &= a_7 f_{span}, & Q_e &= \frac{f_1}{2\sigma_1 |K_1|}, \\ K_0 &= a_3 + j a_4, & Q_0 &= \frac{f_1}{2\sigma_1 (1 - |K_1|)}. \end{aligned} \qquad 6.12$$

Using these relationships and the calculated upper and lower coefficient bounds, useful confidence intervals on these derived parameters can be estimated. If one-off VNA characterisation like this is used to correct for all subsequent multitone system measurements, it is important to characterise what the contribution to measurement uncertainty will be from this calibration. From a single sweep, the calculated 99% confidence bounds turn out to be very optimistic: $f_1 = 249985475 \pm 72$, $Q_e = 2741 \pm 4$ for example. However, a better estimate of repeatability is reached by combining several sets of repeated measurement into one curve fit. Especially if the calibration routine is repeated in between, these fit intervals then become useful indicators of the systematic error on these reference measurements (although definitively no substitute for full systematic error analysis). Three separate repeats were thus made, giving the parameters shown below. These estimates were used as a reference for all subsequent experiments with this resonator.

*Table 6.1: Estimated static resonator constants estimated from calibrated VNA reference measurement*

| Parameter | Value | 99 % fit interval |
|---|---|---|
| $f_1$ | 2.499 805 074 GHz | $\pm$ 380 Hz |
| $\sigma_1$ | 544.554 KHz | $\pm$ 380 Hz |
| $K_0$ | $-0.42922 + j0.72014$ | $\pm 0.015 \pm j0.015$ |
| $K_1$ | $0.00411 + j0.00136$ | $\pm 0.00069 \pm j0.00069$ |
| $Q_1$ | 2295 | $\pm$ 1.6 |
| $Q_e$ | 2738 | $+20/-21$ |
| $Q_0$ | 14200 | $+670/-570$ |

The cavity is over-coupled, with the loaded Q being dominated by the external Q. This is not a problem in practice, although the coupling is not optimum. This coupling regime is better suited to measuring high loss samples rather than for the small perturbation experimental conditions used here. Note the large interval of values for the unloaded Q under this regime. The accuracy of this $Q_0$ estimate is dubious when limited to such a degree by coupling loss. According to the correction for unloaded bandwidth, $\sigma_0 \cong \sigma_1 - \omega_1/2Q_e$, the measured bandwidth is approximately 456 KHz larger than the unloaded bandwidth, which is therefore around 89 KHz.

### 6.1.2 Experimental setup

The prototype SDR system as described in the Hardware chapter was used. Due to the high resonator coupling factor, one RF LNA and a low AWG gain setting (0.07) provided sufficient signal level to best exploit the input range of the ADC under normal operating conditions. At this modulator drive level, the excitation signal distortion was very low. The resonator bandwidth is low compared to the maximum multitone system bandwidth (100 MHz), so a 512-sample excitation waveform consisting of 15 tones covering about 3 MHz and spaced at 195 KHz was used. This was modulated at $f_s/4$, which, as discussed in the Software chapter,

provides excellent rejection of distortion and quadrature errors. Because of this, and the narrow resulting frequency span, predistortion and tone flattening calibration were not required and this step was omitted.

The experiment needed to run in realtime to capture the results from a large number of individual droplets continuously. ADC acquisitions could therefore only be collected at a rate low enough to ensure continuous processing was maintained indefinitely. Due to the bottleneck in the system during the transfer of these acquisitions to the PC, the prototype system was limited to a realtime resonator readout rate of 500 Hz. This is still significantly faster than that possible with a VNA using an optimised, externally triggered configuration. However, of the maximum possible waveform samples that could be acquired, only about 4 % could be used to produce spectral response estimates under these conditions.

Due to these limitations on readout rate, it was not necessary to use an adaptive filter in order to make a continuous spectral estimate. Although the perturbation slew rate caused by the falling drops is still high by the standards of traditional resonator readout systems, simple FFT block processing was still sufficient for estimating the resonator spectral response. Even at the position of highest change, the assumption of stationarity was not noticeably violated over a single waveform capture period. The captured waveform was 16 times longer than the excitation signal, this was then divided into blocks of 512 samples, and the spectral estimates of each block averaged as discussed in the Software chapter. The fast Gauss-Newton algorithm for fitting a Lorentzian response to spectral data was used in order to extract resonant frequency and bandwidth estimates.

### 6.1.3 Results of the falling drop experiments

Several sets of measurements of both ethanol and methanol drop streams were taken. An example of a resonant frequency trace captured for methanol is shown in Figure 6.5 below. The drops appear as short, regularly spaced pulses of resonant frequency shift superimposed on a slowly varying background.

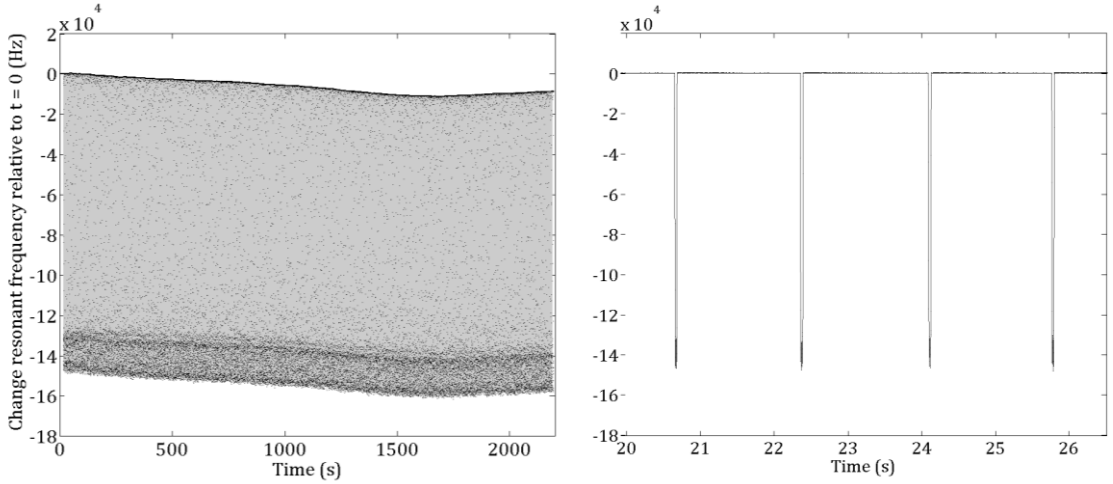



*Figure 6.5: Plot of an entire resonant frequency waveform, consisting of 1 million points over a period of about 30 minutes (left), and a short sample near the start of the experiment (right). 1273 droplets were measured in this time. Measurements are shown as black points joined by a grey line.*

The drift in resonant frequency due to environmental variation, such as thermal cavity expansion, is evident from the baseline position close to zero (seen as a thick dark line at the top of the long waveform in the left plot above). Importantly, the perturbation due to the droplet, shown by the band of points at the bottom of the waveform, appears to follow the drift in 'reference' resonant frequency closely. This demonstrates the potential of dynamic measurements like this to cancel out uncertainties due to drift. As long as the dynamic measurement is arranged so there is a periodic 'reference' signal, like the time between droplets in this case, this can be used as a new reference for the next measurement. This differential technique ensures that drift is cancelled on timescales longer than the recalibration period. To remove the baseline drift in this experiment, a simple moving median filter (MMF) method was used. The MMF implements the statistical median function over a siding block of length 256 samples. Since the signal pulse width is less than 256/2 samples, the MMF was reliable and effective at extracting the baseline drift, which was then subtracted from the main signal.

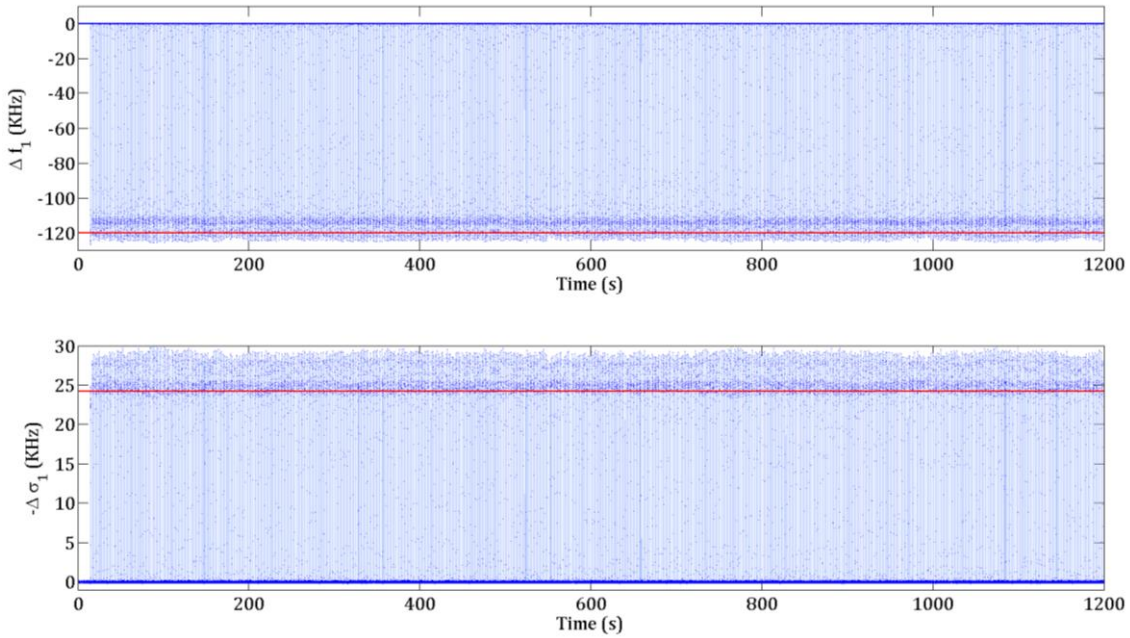



*Figure 6.6: Baseline corrected resonant frequency waveform (top) and bandwidth (bottom) for ethanol droplets. Individual samples are shown by blue points joined by a light blue curve. The red line indicates the perturbation predicted by the spherical dipole theory, based on the sample volume estimated from the average drop rate.*

One of the appealing reasons for working with dynamic signals is the information contained within the time dimension of measurement. The processing of the data captured in this experiment required the detection of each individual droplet, the estimation of the time between droplets (which should be proportional to the droplet volume), and the calculation of the average resonant frequency and bandwidth perturbation per drop. In order to do this a perturbation threshold based on the change in resonant frequency was used as a 'trigger'. The crossing time for this threshold was then linearly interpolated to give an accurate estimate of the drop entry and exit times. The time between these crossing times provides an accurate estimate for the volume of each drop, based on the assumption that the flow rate is constant. The extracted temporal information is summarised in Figure 6.7.

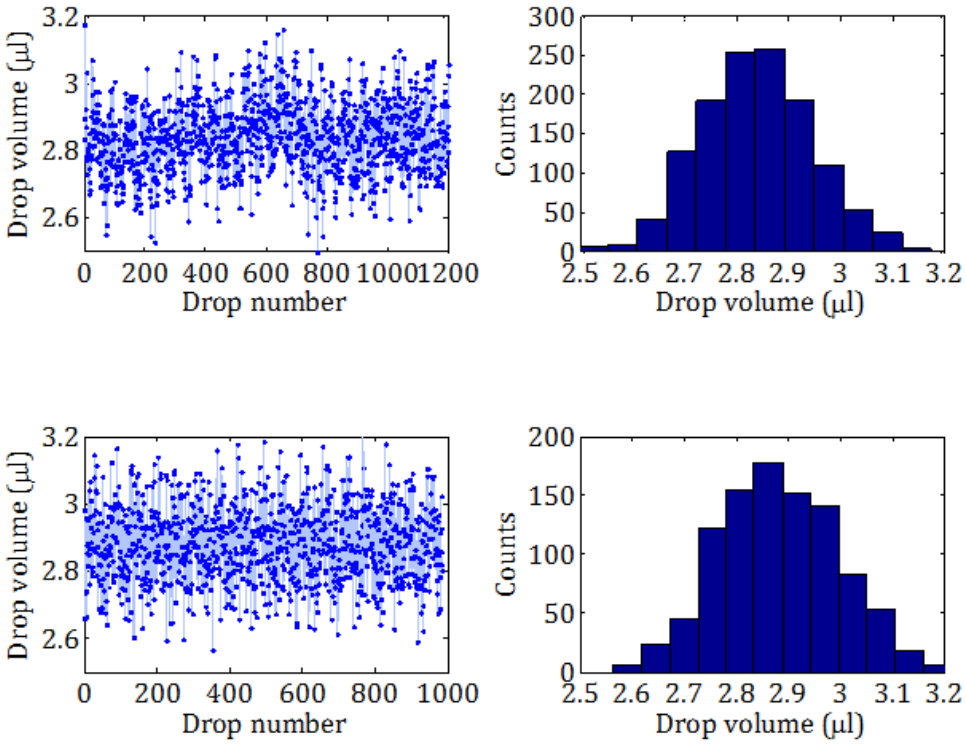



*Figure 6.7: Measurements of drop size dispersion using time-domain measurements of drop interval, top row – methanol, bottom row – ethanol.*

The results seem to confirm that the drop volume is random and normally distributed with mean 2.84 µl for methanol and 2.88 µl for ethanol; a drop diameter of about 1.8 mm. These values are very close; the surface tension of methanol is very close to that of ethanol ($21.68 \cdot 10^{-3}$ Nm$^{-1}$ and $22.31 \cdot 10^{-3}$ Nm$^{-1}$ respectively at 25°C [11]). Using this time-domain method of sample volume estimation in an online measurement system means the inverse equation used can be specific to the every drop – allowing the system to adapt to unknown liquids. This also has the advantage of referencing the dominant source of measurement uncertainty (the sample volume) to the accurate and well-controlled volumetric flow rate, increasing the potential measurement accuracy.

The above stalagmometer equation predicts a drop mass of about 1.1 µg for both alcohols, giving an estimated drop volume of 1.37 µl based on their densities of 0.792 gcm$^{-3}$ and 0.789 gcm$^{-3}$, respectively. This underestimate by just over a factor of 2 is probably due to the imperfect dropper tip – proper stalagmometer

design requires a flat bottom and they are usually made of a hydrophilic material such as glass, rather than hydrophobic PEEK as used here. The drops are bigger than estimated probably because of the initial attraction of the liquid to the outside of the capillary – a behaviour verified by observation. However, an order-of-magnitude agreement still serves to confirm the basic operational principles. It also clearly demonstrates that useful secondary information can be derived from temporal MRS measurement.

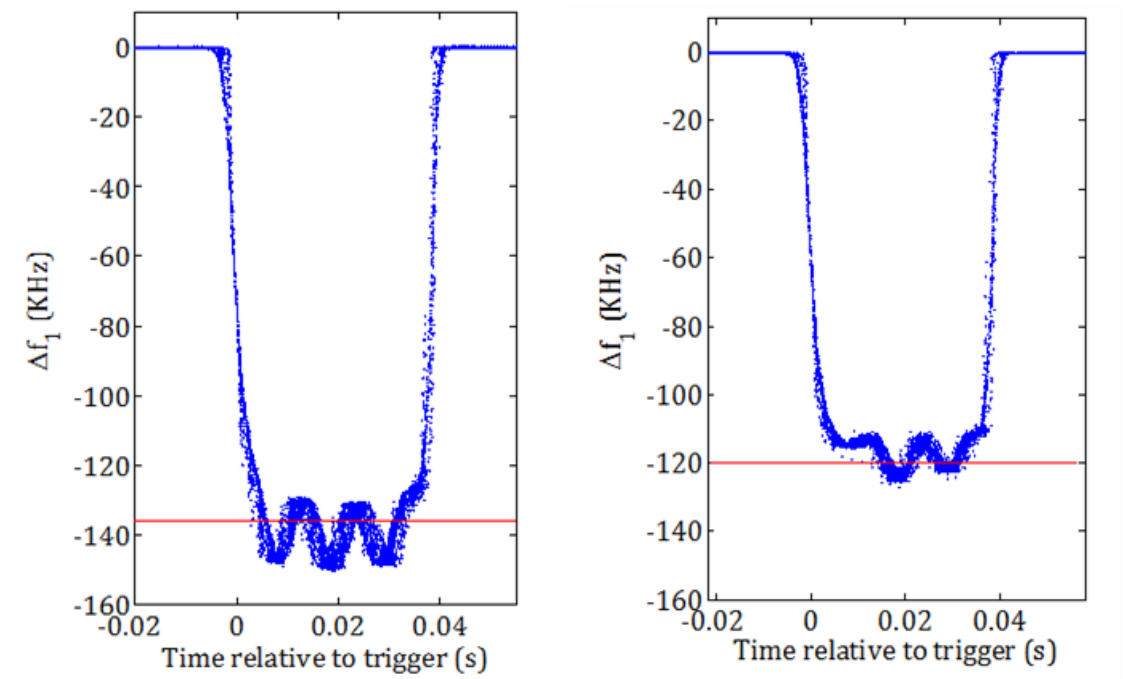



*Figure 6.8: Change in resonant frequency over a large number of repeated drop transients, each plotted as a series of points; left is methanol, right is ethanol. The red line shows the perturbation predicted by theory*

However, despite the promise of this drop method, an apparent problem was found. On processing the individual drop transients, (see Figure 6.8) a significant ripple was observed during the transit of the drop. Initially this was thought to be an artefact of the measurement system – perhaps a 'memory effect' due to the realtime processing algorithm feeding back previous curve fit coefficients, causing the estimates to oscillate around their true values. However, to eliminate this explanation, raw data for the individual tone amplitudes was also taken for off-line

processing. As seen below, the ripple is still present in the raw spectral estimates that have no memory. s

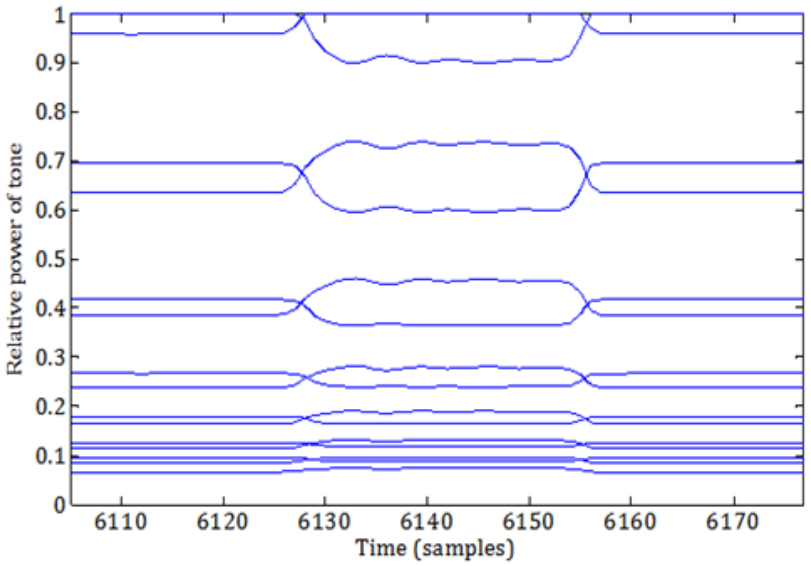



*Figure 6.9: Relative power of the 15 measured tones plotted against time around a typical droplet transient (Sampling rate 250 Hz)*

Therefore, it cannot be an artefact of the curve fitting system. In addition, as described by the time-domain resonator model in the Theory section, the decay time for any transient behaviour due to this perturbation is much less than the sampling interval. This leaves two physical explanations for the ripple effects

- The axial electric field magnitude of the resonator is not homogeneous and instead shows variability with distance – causing variable perturbation as the drop falls
- The droplets are not perfectly spherical when falling and instead oscillate in sphericity due to inertial effects – causing the depolarisation factor to vary as the drop falls

To eliminate the first explanation, a further experiment was done (see Figure 6.10). Chrome steel ball bearings, 1 mm in diameter, were dropped from rest at a height of 13 mm above the resonator directly along the resonator axis. These bearings had very good tolerance for diameter and sphericity, $\pm 2.5$ μm, [12]

making them an excellent standard by which to compare the liquid drops. The theoretical perturbation due to a depolarised good conductor is approximated by taking the limit of Equation 6.8 as $\varepsilon_2 = \sigma/\omega\varepsilon_0 \to \infty$, reducing it to:

$$\frac{p(\theta) - p_0}{p_0} = \frac{-3V_d}{2V_e + 3V_d} \approx -27.70 \pm 0.26 \text{ KHz}. \qquad 6.13$$

The change in bandwidth is predicted to be negligible.

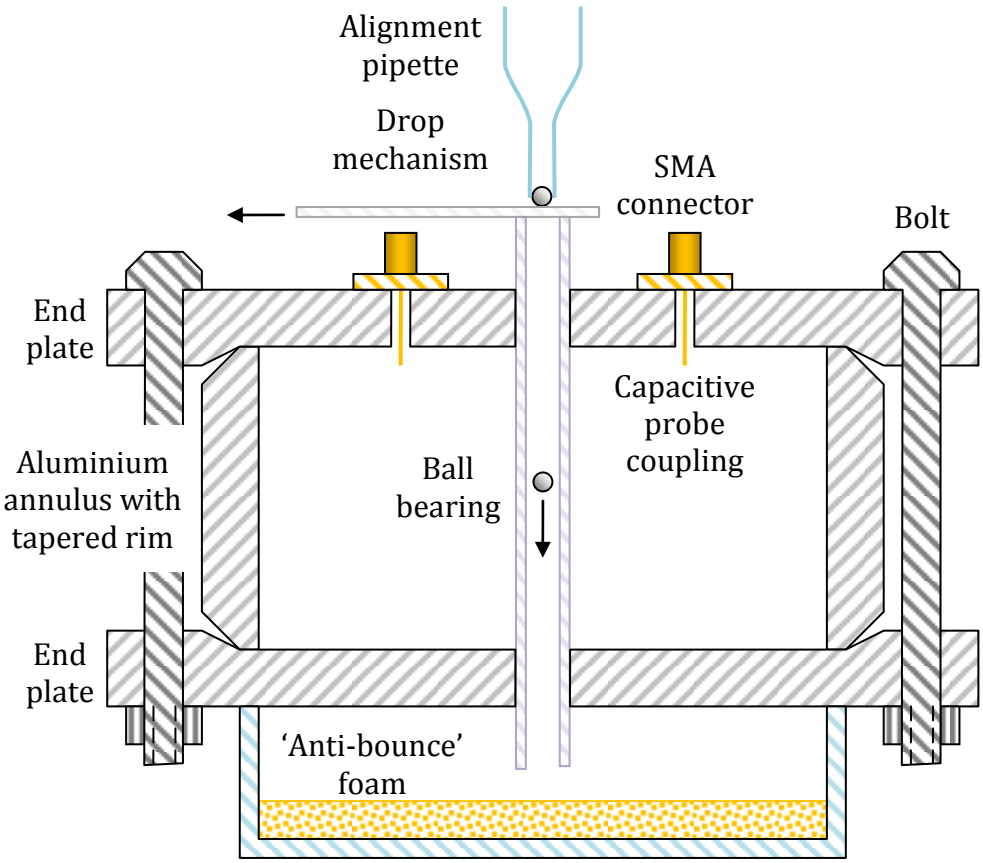



*Figure 6.10: Cross sectional diagram of the $TM_{010}$ resonator configured for the falling ball bearing reference measurement.*

The experiments failed to show any ripple. The results also agree closely with the predicted perturbation, measuring −27.4 KHz at the midpoint of the drop transit, on average. The total width of the perturbation signal is about 50.3 ms. This is higher than the free fall model transit time of 45 ms due to the resonator

perturbation beginning before the particle enters the resonator, and likewise on exit, due to the fringing field around the entry hole.

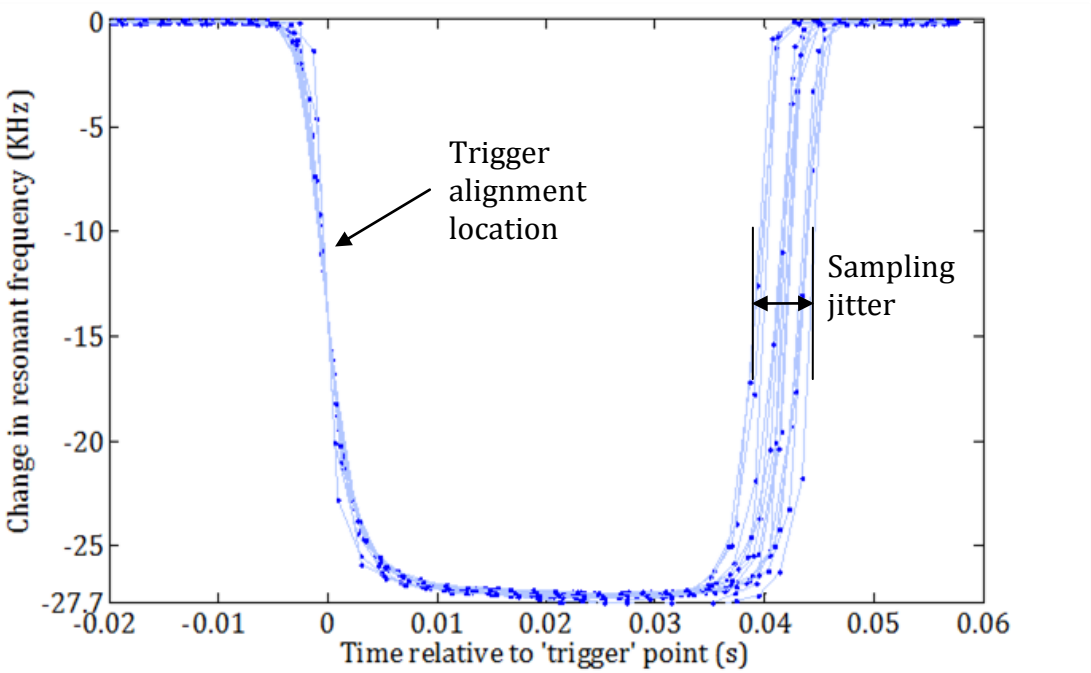



*Figure 6.11: Change in resonant frequency with time over repeated experiments of steel spheres falling from a known height through the resonator.*

There is a relatively large amount of sampling jitter considering the time-domain signal is specifically chosen to be highly repeatable. This may be caused by non-determinism in the LabVIEW software timing and/or digitizer data transfer, occasionally causing a particular trigger to be missed. If a trigger is missed, the sample period defaults to the next trigger – effectively doubling the sampling period for that measurement point. This is consistent with the band of sampling jitter labelled in Figure 6.11, as traces appear to be discreetised into bands of an integer number of samples. To investigate this explanation, software timing was used to measure the time taken to fetch successive acquisitions from the digitizer. A fetch period exceeding twice the desired sampling period would imply a missed trigger and hence explain this measurement artefact.

There is no truly accurate high-precision timing available in a multithreaded operating system environment like Microsoft Windows®. The best that LabVIEW offers is the millisecond timer, which returns the current number of milliseconds from some unspecified time (possibly since last boot-up). However, a higher precision timer is available in the Windows application-programming interface (API) that is capable of measuring smaller intervals ($< 1\mu s$ on the computer running the experiments). This API function was used in the LabVIEW control software to measure the acquisition period (Figure 6.12). Whilst the majority of the samples appear to be very close to the desired 2 ms period – implying that the right trigger has been received and the timing is accurate – occasional acquisitions fail to meet their deadline and result in a missed trigger. Whilst this particular example has only an extremely small probability of a missed trigger (just 0.01 %), this is likely to be dependent on the average background process load of the computer and similar uncontrolled factors. This could therefore explain the observed sampling jitter on the resonant frequency waveforms.

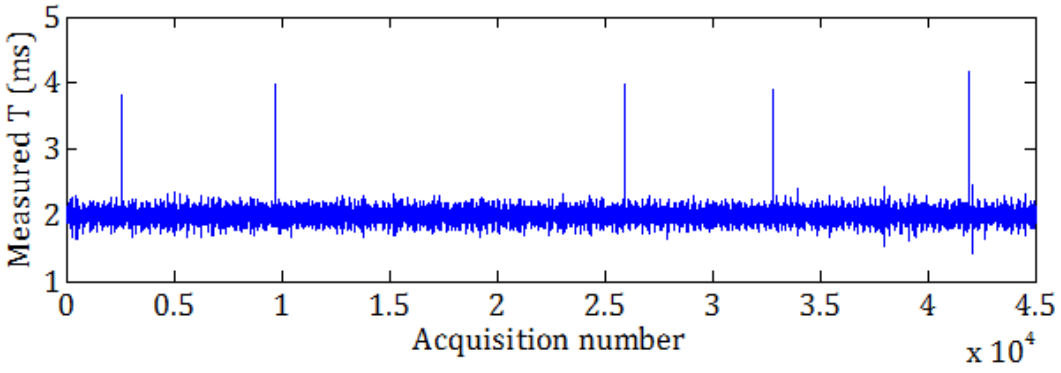



*Figure 6.12: Software timed acquisitions showing occasional missed triggers – a possible cause of timing jitter.*

The non-deterministic software environment of the PC is therefore causing a measurement artefact, however the hardware system should be fully deterministic and is coordinated accurately with a shared 100 MHz clock, itself frequency locked to a precision 10 MHz reference in the RF signal generator. The long-term solution to sampling jitter is therefore to perform the digital signal processing on a fully deterministic system – such as a realtime operating system, embedded processor or, most desirably, an FPGA. This would not only allow for precision timing with

minimal sampling jitter but also allow for 100 % utilisation of the available waveform data. This, in turn, promises a much higher maximum sample rate and increased dynamic range, as mentioned previously in the Hardware chapter.

### 6.1.4 Drop oscillations

Having ruled out measurement-related sources of the observed oscillations, a physical explanation was sought. It turns out that falling droplets naturally undergo surface oscillations (mechanical resonance) due the second order system created by the drop surface energy (spring), inertia (mass) and viscosity (damper) [13]. Whilst the alcohol drops have a low mass, they also have low viscosity and high surface energy because of hydrogen bonding; hence, drop oscillations are still plausible. The axisymmetric resonant modes of the liquid drop boundary were first described by Lord Rayleigh in 1879, [14], [15]:

$$F(\phi, t) = r[1 + \epsilon \cos \omega t P_n(\phi)], \tag{6.14}$$

where $\phi$ is the angle measured from the centre of a drop of radius $r$, and $P_n(\phi)$ is a Legendre polynomial for the resonant mode $n = 2,3, \ldots$. The resonant frequency of these surface oscillation modes is given by

$$\omega^2 = \frac{\gamma}{\rho r^3} n(n-1)(n+2). \tag{6.15}$$

This predicts a resonant frequency for fundamental vibrational mode of the drops of 91.8 Hz for methanol and 90.6 Hz for ethanol. This is a high frequency, but with the system sampling at 500 Hz, it should be possible to check whether the observed ripple matches it.

As the number of points in the oscillatory part of each drop is relatively few, a nonlinear least squares algorithm was used to fit a sinusoidal wave to this part of every drop waveform; *i.e.,* $\Delta f_1(t) = a_1 + a_2 \cos a_4 t + a_3 \sin a_4 t$. The fitted frequency was estimated for every drop. A histogram of the measured drop oscillation frequency is shown in Figure 6.13. An R-square value of 0.99 or higher

was obtained for the majority of the droplets, showing that the sinusoidal model agrees very well with the measured data. This also implies that any damping of the oscillations is small enough not to cause appreciable decay of the oscillations during their transit time. Furthermore, if higher order droplet vibrational modes are excited, either they are of insignificant amplitude or they do not affect the depolarisation factor in such a way as to cause resonator perturbation oscillations.

This analysis shows that the majority of the measured drops have an oscillation frequency close to that predicted by the Rayleigh model, perhaps biased towards a slightly lower resonant frequency than predicted (Figure 6.13). However, this is still a good enough agreement to conclude that this natural behaviour is the cause of the observed ripple in resonant frequency. That the amplitude of the oscillations is so high is remarkable, testament to the sensitivity of depolarisation factor to shape. It also seems somewhat counter-intuitive to the accepted wisdom that inertial effects are insignificant on these length scales. Clearly, this assumption can be easily mistaken in practice.

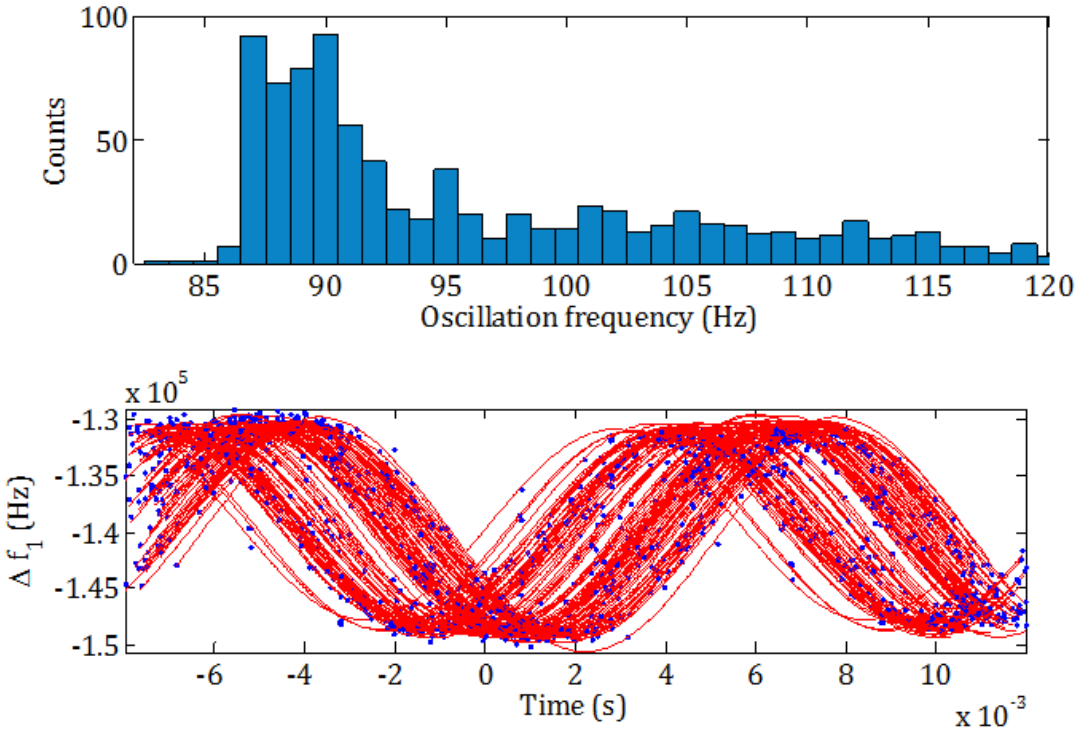



*Figure 6.13: Histogram of measured drop oscillation frequency derived from time-domain resonant frequency waveform processing of individual droplets (top) and example of the Fourier curve fits to measured droplet oscillations (bottom) with fit as red curve and data points as blue points*

## 6.2 Measuring a gas segmented flow

To provide a more controlled perturbation signal, and to better evaluate the full capabilities of the multitone system under large-signal perturbation conditions, a second experiment was devised. An existing split ring resonator (SRR) [16], a compact lumped element resonant sensor with a highly localised electric field, was used to measure the perturbation caused by a capillary containing an alternating flow stream of a pure reference solvent with air. This sensor has an extremely high filling factor, as its electric field is highly confined. The resulting high sensitivity means the perturbation signal caused by even low permittivity samples is exceptionally large, with a resonant frequency slew rate of around 1.9 GHz/s.

Compressed air and solvent flow from a syringe pump were combined at a capillary tee junction to generate this flow stream. As discussed in the Applications chapter, on small length scales such as this junction a laminar flow regime and a dominance of surface forces exist, leading to interesting behaviour such as highly stable multiphase flow regimes. This is exploited here to create regular segments, or 'slugs', of air interspersed with solvent – rather than small, irregular, isolated bubbles expected on conventional scales. This flow regime is analysed in detail by Triplett *et al.* [17][18]. Gas-liquid segmented flow like this is often used to enhance the accuracy of chromatographic assays by reducing dispersion, [19] to enhance mixing in each phase, [20] and to perform gas-liquid reactions and separations [21]. The constantly refreshing gas-liquid interface also enhances the reaction rate of chemical reactions, and allows for the extraction products continuously, which enhances yield in microfluidic reactors.

As the air segments extended across the entire capillary diameter, and were long enough to occupy the whole electric field sensing region of the SRR, they created a momentary 'zero reference' point where the capillary nominally contained only air. This was then followed by a long slug of solvent, and the combined segmented flow stream passed the split ring sensor at a velocity of about 20 mm/s. A suitable, large amplitude, dynamic perturbation signal with which to test the multitone system could thus be generated.

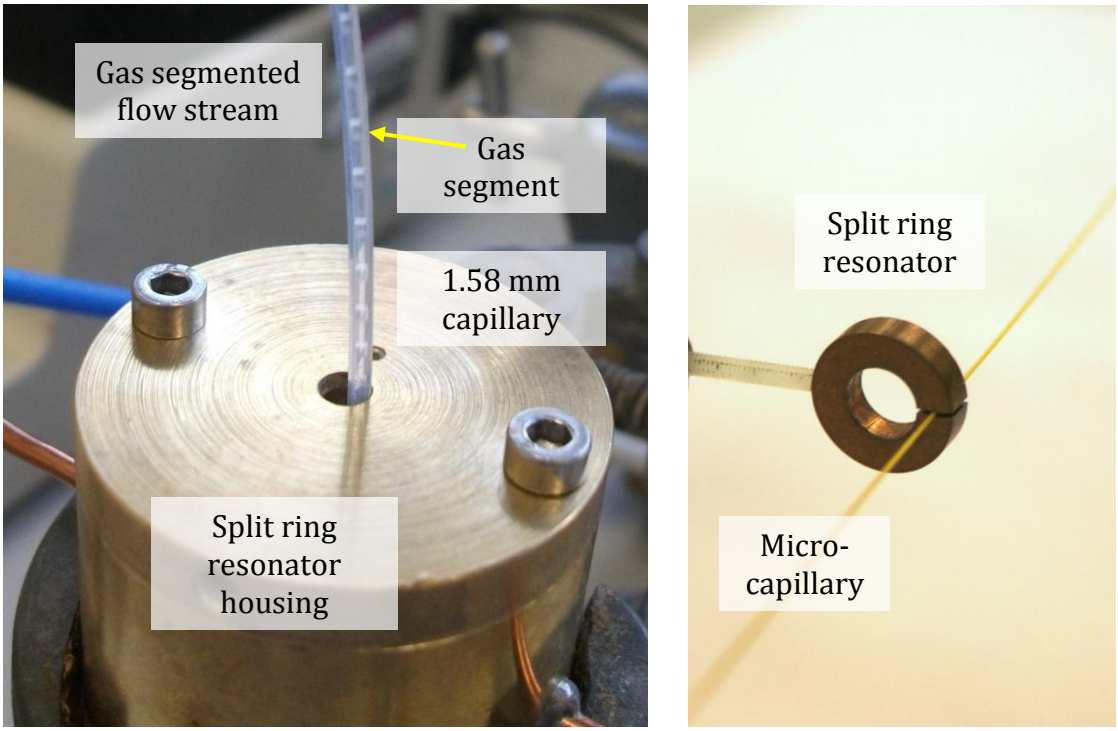



*Figure 6.14: Photograph of the split ring resonator housing with inflowing gas segmented flow stream (left) and the split ring resonator with fitted with microcapillary (right).*

Due to the high filling factor of the SRR, high dielectric loss liquids such as water, methanol, *etc.*, cause a massive increase in bandwidth, making measurement of the resonance difficult and inaccurate. The SRR is better suited to measuring small amounts of polar contaminant or analyte in a largely non-polar matrix. Such compositional analysis was its intended application. This meant that a suitable low loss reference liquid was needed for characterisation. Toluene ($C_6H_5CH_3$), a benzene ring with single methyl group, is a common solvent used in industrial and pharmaceutical processes. It is only very weakly polar, has low viscosity yet a relatively high boiling point, making it an ideal candidate. According to Petro and Smythe, [22] toluene has static dielectric constant of $\varepsilon_s = 2.3837$, and the first dielectric relaxation at a frequency of 24.57 GHz, with an associated dielectric increment of only $\Delta\varepsilon = 0.118$. This gives a complex permittivity at the resonant frequency of 3.58 GHz of about $\varepsilon_2 = 2.38 - j0.017$. This very low dielectric loss, $\tan\delta = 0.007$, is guaranteed not to 'swamp' the resonant sensor.

### 6.2.1 Experimental setup

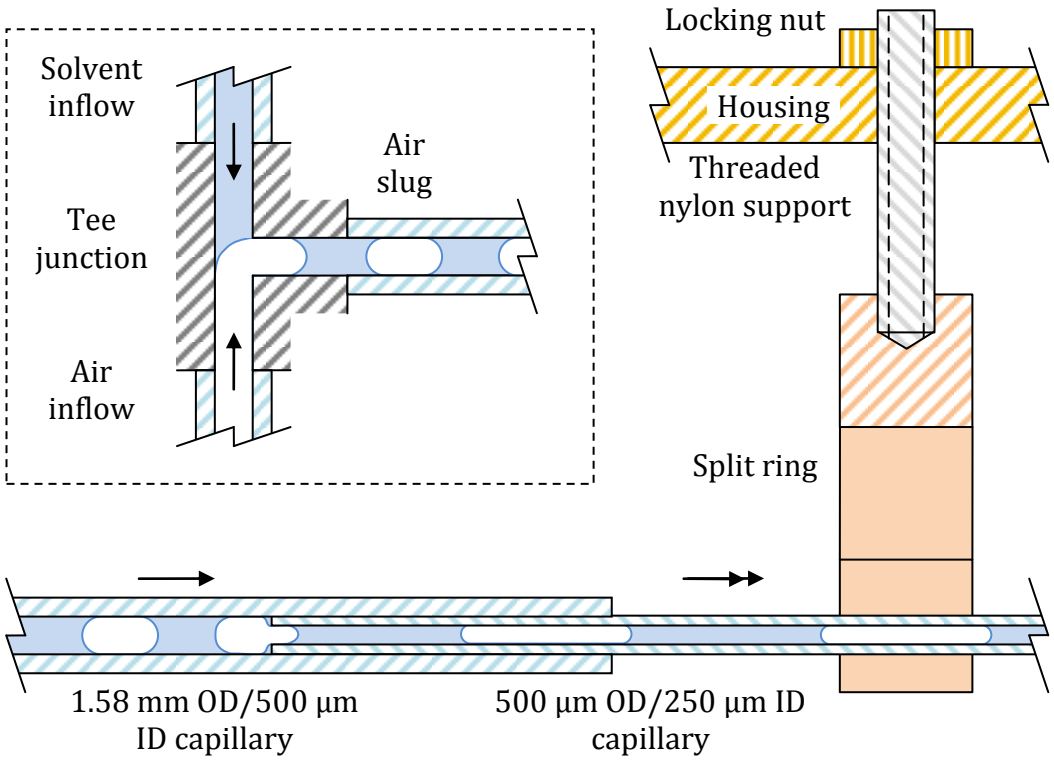



*Figure 6.15: Cross sectional diagram of the gas segmented flow setup showing the split ring resonator and capillary configuration – Inset: generation of gas segmented flow at a tee junction*

To create a resonator perturbation that alternated between two known static values, the gas segmented flow was eluted into a 500 µm OD/250 µm ID Teflon® AF capillary (Biogeneral, CA, USA) before passing through the capacitive gap of the split ring. This had the effect of stretching the slugs by a factor of four in length as the narrower capillary had half the inner radius. The longer slug completely filled the active length of the capacitive gap – for a split second being equivalent to homogeneous flow conditions. The outer diameter of this capillary was conveniently sized so that a seal was possible simply with a push fit into the larger capillary. Teflon® AF (grade 2400) is an amorphous fluoropolymer that has one of the lowest known dielectric constants of any solid and a very low loss factor, $\varepsilon_1 \approx 1.89 - j3 \cdot 10^{-3}$, [23], having minimal impact on the measurement accuracy and $Q$ factor. Pumping a fluid through a narrow capillary can require large forces,

so the length of this narrow capillary was kept short in order that the syringe pump pusher mechanism could cope and gas pressure was required was reasonable. The join was located inside the resonator housing. The fluid eluted into a containment trap, sealed using a conformable polymer film (Nescofilm®, Japan). A second, large diameter length of tubing extending out of this trap allowed displaced gas to escape to atmospheric pressure inside a fume cupboard – this was important as toluene produces vapours harmful to human health [24].

The controlled gas flow was created by tapping-off the laboratory compressed air supply (from an external compressor) using a manually controlled adjustable air bleed and pressure meter readout. A pressure of about 5.6 KPa was found to be sufficient to create a slug flow regime with the syringe pump flow rate set to 200 µl/min; depending on the length of the narrow capillary. Experimentally, the optimum pressure was found by releasing the constriction until the point at which liquid was seen moving back along the gas line. The pressure was then gradually increased until it was sufficient to force the liquid slowly back towards the tee junction, at which point relatively stable slug flow could be created and adjusted for gas-liquid ratio by fine adjustment of the pressure. Although simple, this method of generating a constant pressure gas source is not very accurate, and as the results show, leads to quite a large variation in the gas segment length. This was not an issue for this experiment, but a better and more repeatable experiment would be achieved using bottled, compressed nitrogen.

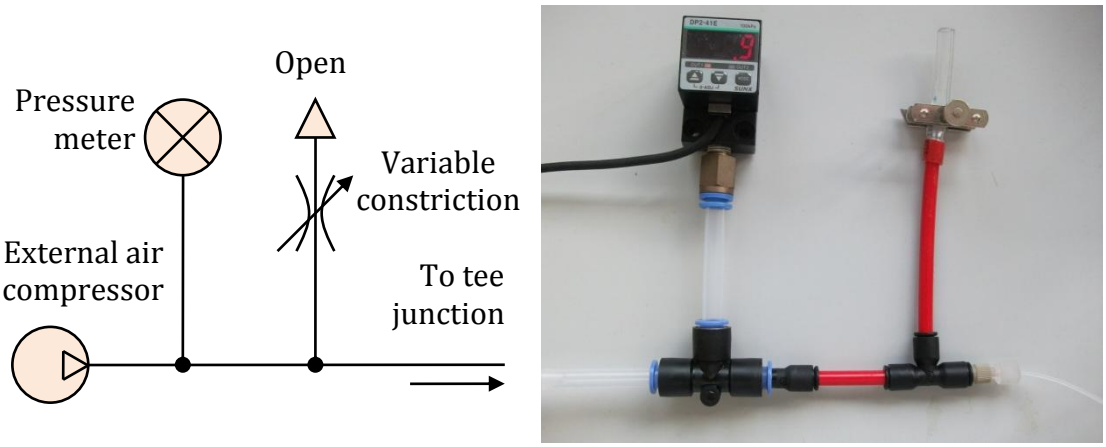



*Figure 6.16: Configuration of the air supply (left) and photograph (right)*

In order to increase the dynamic range of the measurement for the multitone system, the coupling to the SRR was increased by moving the coupling loops inwards. The brass housing was also re-polished and the copper ring was cleaned by sonication in an acetone bath to remove traces of metal polish, fingerprints, *etc.* Subsequent handling of the split ring was avoided. Having a high sensitivity to evanescent electric fields at the surface of the copper ring, this cleaning procedure is important to maximise the quality factor of the sensor. Other than this, no alterations were made to the resonator, the design and construction of which is covered in more detail in [16].

### 6.2.2 Resonant sensor characterisation

The analytical perturbation analysis for the split ring is similar to that of the droplet experiment, as the sample is depolarised in the SRR too. The dipole energy term in the general perturbation equation is therefore convenient here also. The dipole energy of a capillary containing a sample is a bit more complex than a homogeneous sphere: [16]

$$e_j \Delta \tilde{\varepsilon}_j \approx 2V_t \cdot \varepsilon_0 E_0^2 \cdot \frac{(\varepsilon_1+\varepsilon_2)(\varepsilon_1-1)+(\varepsilon_2-\varepsilon_1)(\varepsilon_1+1)(a/b)^2}{(\varepsilon_1+\varepsilon_2)(\varepsilon_1+1)+(\varepsilon_2-\varepsilon_1)(\varepsilon_1-1)(a/b)^2}, \qquad 6.16$$

where $a$ is the capillary inner radius, $b$ the outer radius $\varepsilon_2$ the sample complex relative permittivity and $\varepsilon_1$ the capillary relative permittivity. $V_t = \pi a^2 \ell$ is the total

volume of fluid within the electric field region of the split ring (assumed to be uniform over this length). Note the use of $a$ rather than $b$ in the equation for $V_t$ is a correction to that quoted in [16]. This is because the volume integral of the energy density is over the sample inside the capillary, rather than over the entire capillary volume itself, and, since the field inside a depolarised ellipsoidal prism or body of revolution is always uniform, this volume integral reduces to the geometric volume multiplied by the constant internal field. This result is derived for a dipole in an infinite expanse of free space. However, the top and bottom faces of the split ring capacitor are close enough to be in physical contact with the capillary, which will cause a degree of error. Whilst [25] gives a correction for the proximity of conductors based on image theory, the correction is in terms of effective depolarisation factor and thus is difficult to apply to the above equation.

A simple approximation to find the electric field energy term, $e_i$ of the unperturbed resonator is to use the ideal parallel plate capacitor assumption, which states that the electric field in the gap is uniform within the gap and zero everywhere else – an assumption that is clearly invalidated to some degree by inspection of the FEM simulation (see Figure 6.18). This is likely to cause this theory to overestimate the magnitude of the perturbation, as the fringing fields will tend to increase the true effective volume. Under this assumption, the effective volume $V_e$ of this uniform field is therefore simply that of the air cuboid in the gap. Thus, in the static field limit, $e_i = h_i = \varepsilon_0 E_0^2 V_e$ and, with $\Delta\widetilde{\varepsilon}(\varepsilon_1, \varepsilon_2, a, b)$ standing in for the complicated fraction in the equation above, the relative eigenvalue perturbation is given by:

$$\frac{p(\theta) - p_0}{p_0} = \frac{-V_t \cdot \Delta\widetilde{\varepsilon}(\varepsilon_1, \varepsilon_2, a, b)}{V_e + V_t \cdot \Delta\widetilde{\varepsilon}(\varepsilon_1, \varepsilon_2, a, b)}. \qquad 6.17$$

The results of this simple perturbation model are shown below for various sample relative permittivity values. This simulation is based on the measured resonant frequency and loaded bandwidth, nominal gap width of 500 μm, split ring radii of 6 mm and 3 mm, split ring length of 3 mm, capillary OD of 500 μm and capillary ID of 250 μm. For the reference permittivity of toluene, this equation predicts a change in resonant frequency of $-42$ MHz and bandwidth increase of 330 KHz.

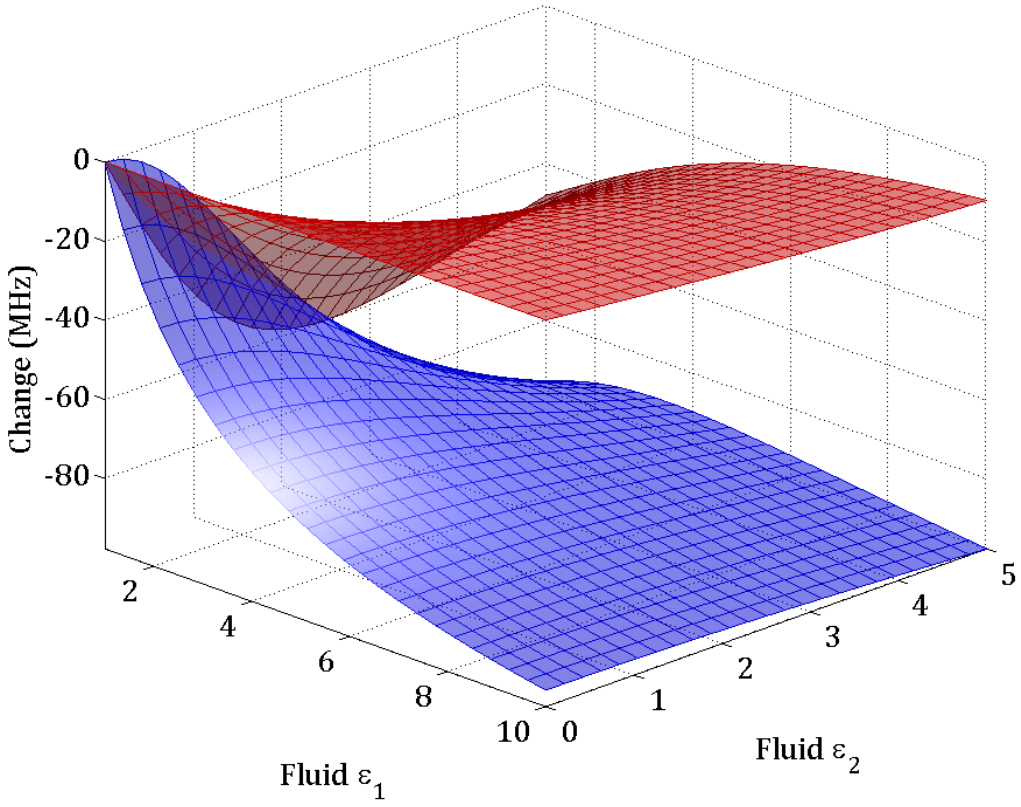



*Figure 6.17: Predicted change in $f_1$ (blue) and $\sigma_1$ (red) against sample permittivity according to Equation 6.17*

The nature of the electric fields in the gap region with capillary is shown in Figure 6.18, from a 3D eigenvalue simulation of the entire SRR in COMSOL. In order to reduce the computational burden, symmetry boundary conditions were exploited in the equatorial and vertical planes (perfect magnetic conductor and perfect electric conductor, respectively). Quartering the volume in this way reduced the computational burden significantly. The final meshed model had 840,000 degrees of freedom and took 4 minutes to solve using a 64-bit, 12-core workstation with 12 GB of RAM. The field plots shown are for an air-filled capillary. With air, some enhancement of the internal field within the capillary can be seen. This effect occurs when the permittivity of the capillary exceeds that of the fill – something that could be exploited for a sensitivity enhancement if sapphire or other high dielectric constant capillaries were used. The field magnitude within the capillary is still quite uniform, despite the proximity of the conductors. The effect of fringing field can also be seen as a 'glow' emanating from around the edges of the gap. Comparing two eigenvalue simulations, one with toluene and one with air, this

FEM model predicts a change in resonant frequency and bandwidth of −36.00 MHz and 336.2 KHz, respectively.

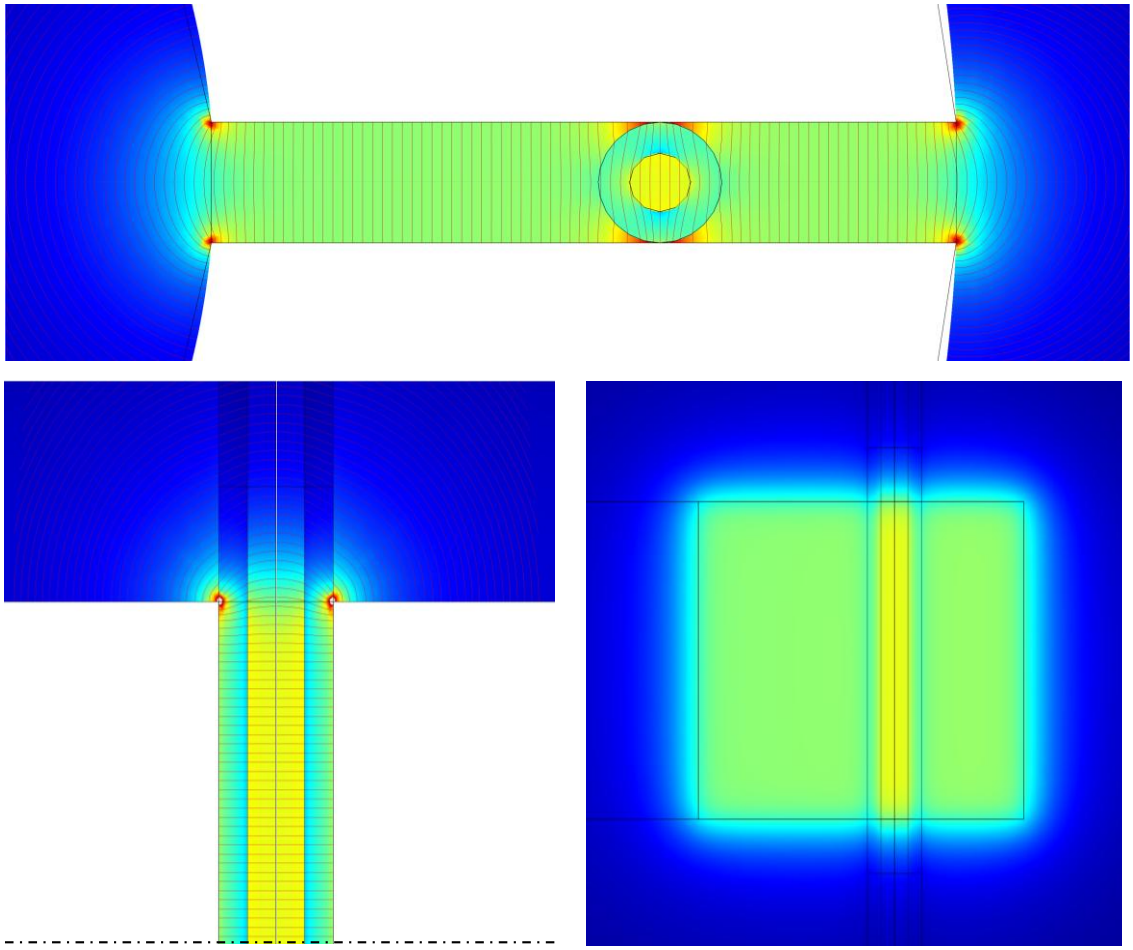


*Figure 6.18: FEM simulation of the SRR with air filled capillary. |E| is shown by shading blue-red, E field lines by red curves. The top image is a section through the ring equatorial plane, bottom left is a vertical section through the capillary, and bottom right is a section through the middle of the split; the E field direction being normal to this image*

Parametric eigenvalue perturbation measurements were performed over a sample permittivity grid. A 21 × 21 grid was used, with identical mesh and geometry to the simulation shown above in Figure 6.18. The total simulation time was approximately 25 hours. Figure 6.19 shows the resulting two-dimensional inverse function – giving complex sample permittivity as a function of change in bandwidth and resonant frequency. These data could be interpolated using a radial basis

function or cubic spline if this function needed to be estimated in an online sensor system.

It was found that such direct, non-parametric interpolation of the simulated eigenvalue perturbation was much more accurate than the traditional method of using FEM modelling simply to extract and estimate of the volumetric field integrals, and thus to merely derive static estimates for the effective volumes of the resonator and sample. This is because field integrals derived from FEM require an additional numerical estimation step, making them more sensitive to the effects of finite mesh quantisation. Calculating the difference in eigenvalues, on the other hand, tends to cancel out quantisation errors, especially if the mesh is kept constant as the sample permittivity is varied, as done here. In fact, being a differential comparisons technique, it is actually very analogous to real-life resonator perturbation, but instead of cancelling drift, it instead partially cancels mesh quantisation error.

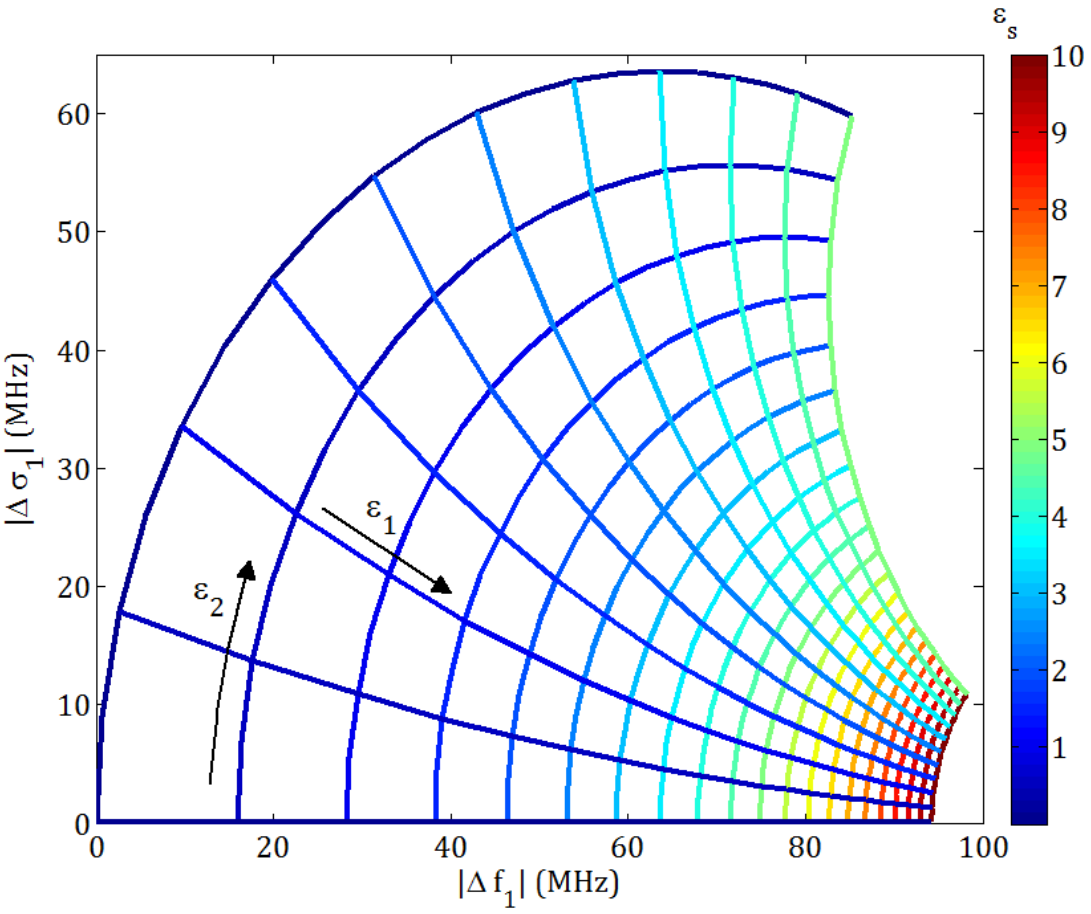



*Figure 6.19: Complex permittivity inversion nomogram derived from parametric FEM simulation of a split ring resonant sensor. Lines are spaced at relative permittivity steps of 0.5.*

The error between the FEM prediction and the analytical prediction, $e \triangleq \Delta\hat{p}_{1,FEM} - \Delta\hat{p}_{1,EQN}$, is shown in Figure 6.20. The disagreement is quite high. When $\varepsilon_1$ is low the analytical equation overestimates the resonant frequency perturbation by as much as 25 %, and when $\varepsilon_1$ is high for a low-loss sample, the bandwidth is underestimated by as much as 42 %. By inspection of the error function, it is also possible to conclude that a simple compensation to the effective volume is not sufficient to reduce this error significantly; therefore, the source of this discrepancy must be due to sample-dependent field effects – *i.e.,* non-linear perturbation.

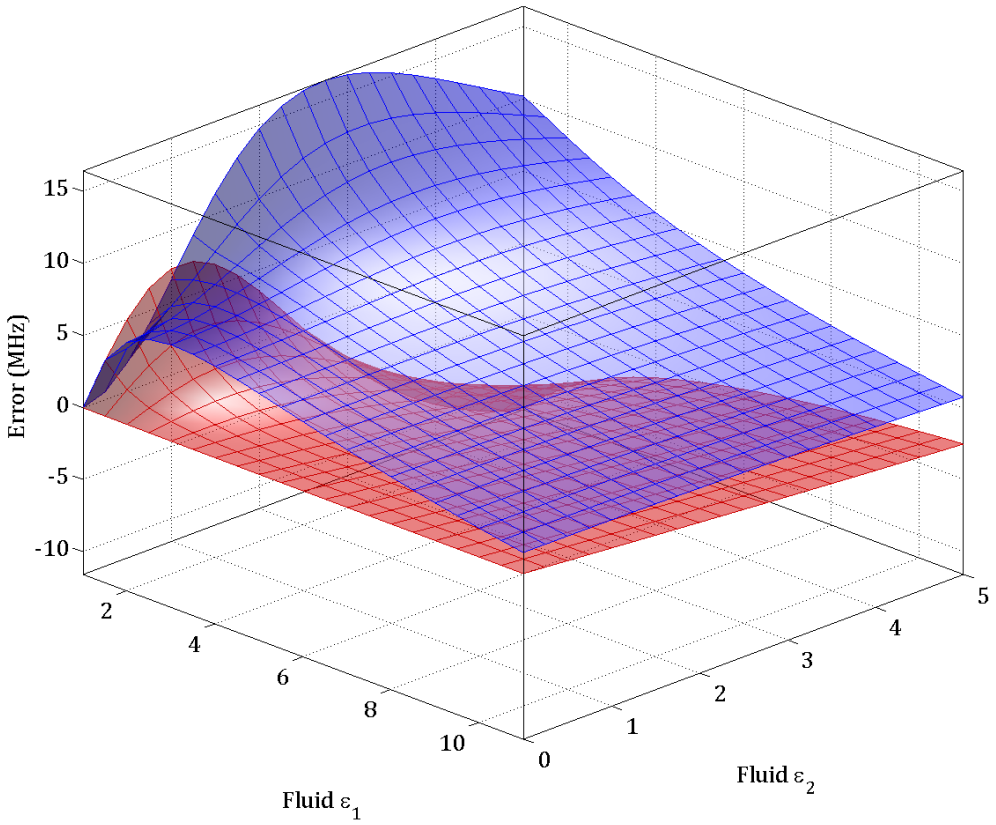



*Figure 6.20: Comparison of the analytical prediction via Equation 6.17 and the parametric FEM simulation. The error in the predicted change in $f_1$ is shown in blue and that in $\sigma_1$ in red, against the 2D sample permittivity grid.*

The FEM simulation should give a much more accurate estimate of the perturbation than the simple dipole theory. However, as microscope inspection of the split ring reveals (see below), the geometry is poorly controlled, with the split ring gap being neither parallel nor uniform in width, and the central duct in the capillary being somewhat off-centre. Because of these physical imperfections, the FEM simulations were unrepresentative and will be subject to error.

*Figure 6.21: Photomicrograph of the gap section of the split ring resonator showing a cleaved section of Teflon® AF microcapillary*

Due to these inherent inaccuracies of both analytical and numerical modelling approaches, it was decided to opt for a straightforward comparative study to verify this experiment. In other words, whilst this may be a difficult sensor to model and invert accurately, as long as the multitone system results agree with static measurements made by the VNA then this still serves the overall experiment aim of system verification. If this sensor design were to be used in a real application then the fabrication process would need to be improved, and a more dimensionally accurate capillary; *e.g.,* precision-made fused quartz, would need to be employed to guarantee accuracy around the 1 % level typical of well-designed MRSs, rather than 10-20 % as it is currently.

As with the previous experiment, comparative calibrated VNA measurements were taken for the SRR with the capillary being air-filled and after static filling with toluene. Four separate measurements were taken, and the spread of values used to set approximate confidence intervals on these measurements.

*Table 6.2: Estimated static SRR constants from multiple calibrated VNA reference measurements.*

| Parameter | Value | 99 % interval |
|---|---|---|
| $f_1$ | 3.580 767 667 GHz | $\pm$ 1.500 KHz |
| $\sigma_1$ | 1.506 423 MHz | $\pm$ 1.500 KHz |
| $K_0$ | $-0.0067 - j0.0047$ | $\pm 0.000\,12 \pm j0.000\,12$ |
| $K_1$ | $0.44 - j0.23$ | $\pm 0.0027 \pm j0.0020$ |
| $Q_1$ | 1189 | $\pm$ 1.2 |
| $Q_e$ | 2365 | $\pm$ 9 |
| $Q_0$ | 2389 | $\pm$ 14 |

*Table 6.3: Perturbation constants for toluene under static conditions from calibrated VNA reference measurements.*

| Parameter | Value | 99 % interval |
|---|---|---|
| $\Delta f_1$ | $-37.237$ MHz | $\pm$ 12 KHz |
| $\Delta\sigma_1$ | $+315.6$ KHz | $\pm$ 12 KHz |

The static change in resonant frequency and bandwidth are reasonably close to the $-42$ MHz and 330 KHz predicted by the above perturbation theory, especially considering the geometric imperfections, neglected conductor proximity, and fringing field effects. They are even closer to the $-36$ MHz and 336 KHz predicted by the numerical modelling, suggesting that parametric FEM simulation is a more accurate inversion method for this resonator despite its physical unrepresentiveness to the geometric imperfections of the real sensor.

Dealing with these rapid, large shifts in resonant frequency was a demanding test of the multitone system. A full-span multitone excitation of 64 samples was used, resulting in 31 tones spaced to avoid DC and sideband images. However, with a tone spacing of 3.125 MHz – comparable to the 3 dB bandwidth of the resonator – only about 3-4 of these tones were significant at any given time in determining the

resonator parameters due to the statistical weighting in the curve fit algorithm. As in the previous experiment, FFT-based spectral estimation with 128 averages was used, combined with simple fast Lorentzian curve fitting. A sampling rate of 500 Hz was possible under these settings. Full nonlinear predistortion correction was found to be essential, and a calibration grid of 16 × 16 amplitudes was used followed by tone amplitude equalisation – as discussed in the Section 5.2.1.

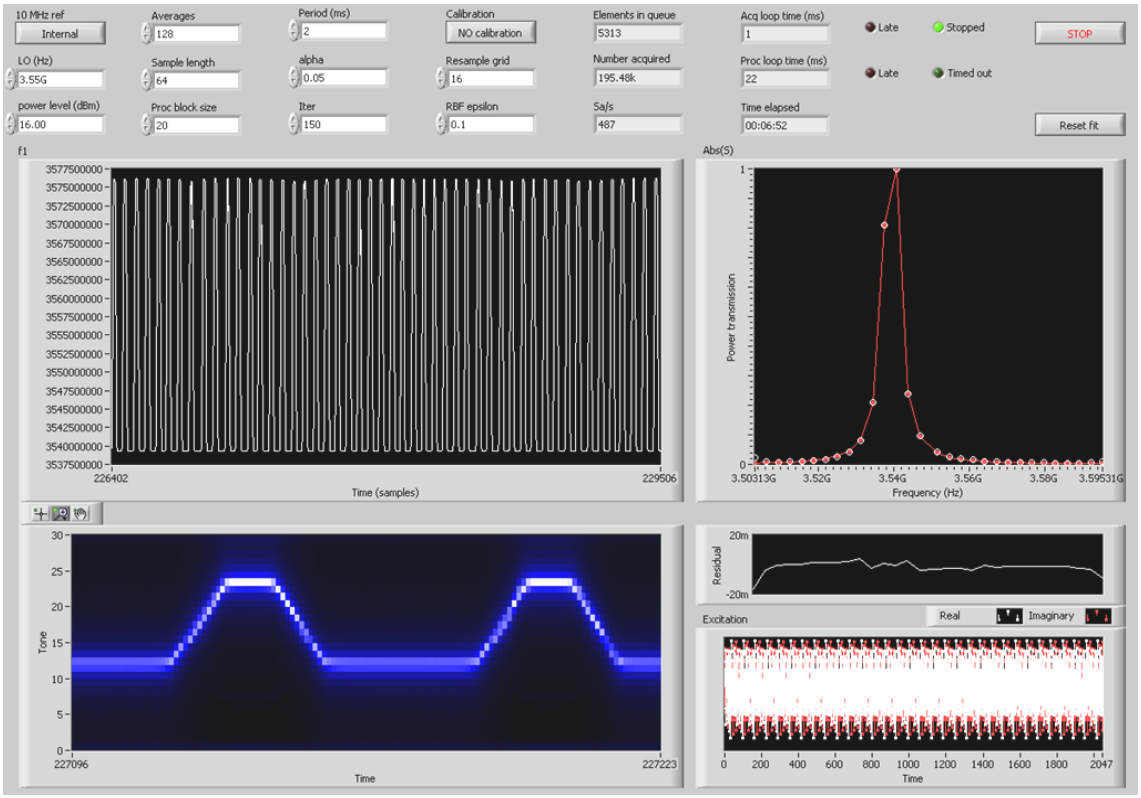



*Figure 6.22: Screen shot of the multitone system LabVIEW front panel UI with the gas-segmented flow experiment in progress. Clockwise from top left, the graphs show resonant frequency vs. time, spectral power transmission estimate and curve fit, fit residual vs. frequency, the excitation signal vs. time and a spectrogram showing the power transmission as a tone gradient black-blue-white.*

### 6.2.3 Results of the gas segmented flow experiment

Typical time-domain segmented flow measurement waveforms are shown in Figure 6.23.

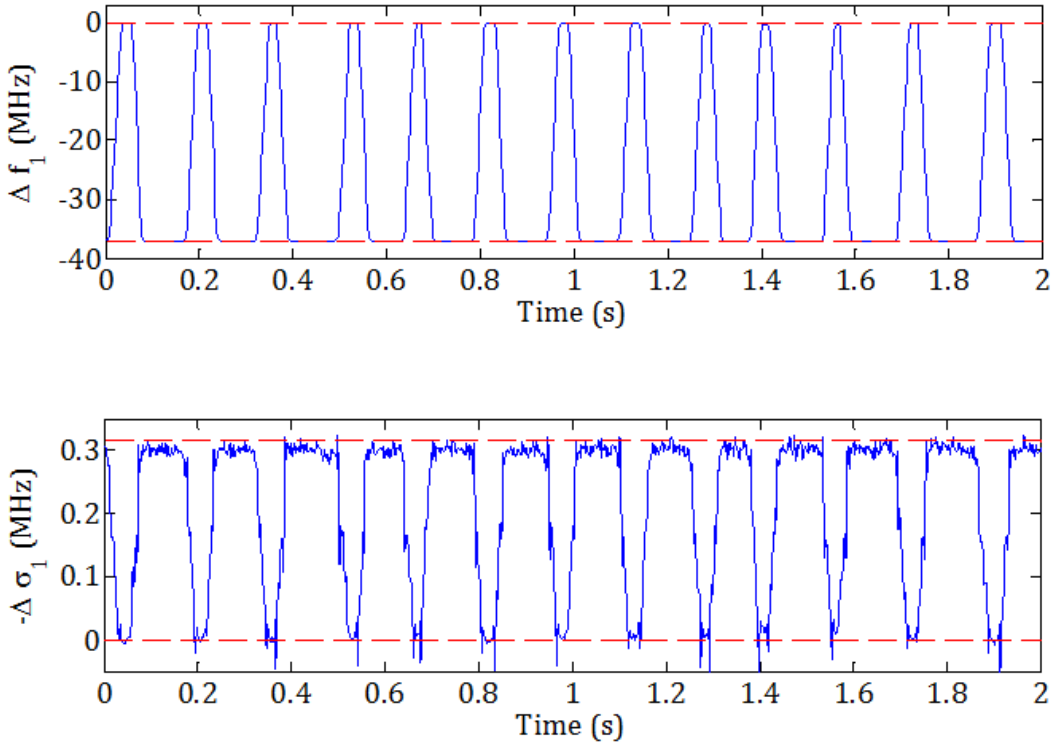



*Figure 6.23: Measured change in resonant frequency (top) and bandwidth (bottom) over the first 2 seconds of the experiment measuring toluene-air segmented flow; red dashed lines show the expected perturbation limits corresponding to the static VNA measurements.*

The resonant frequency measurement is generally good, and agrees quite closely with the static VNA measurement, showing an average resonant frequency change of −36.959 MHz, only 0.7% less that that measured with the VNA. Such a small discrepancy could easily be due to small differences in the sensor, or evaporation of the toluene into the gas segment – increasing its permittivity relative to air. The bandwidth, however, is not such a good measurement. It shows a disproportionate amount of variation, overshoot on the gas segment cycle and a systematic glitch around the midpoint of the transition. These errors most likely caused by the tone spacing being too large to allow a reliable

fit for bandwidth. However, moving to a 128-sample, 61 tone waveform, which would largely eliminate these errors, would have significantly reduced the amplitude of the signal at the receiver, leading to degraded precision on the resonant frequency measurement. Further processing was therefore only applied to the resonant frequency waveform.

*A similar time-domain analysis was performed as with the droplet trigger signal was generated by applying a threshold to the resonant waveform in order to isolate the rising and falling edges of the alternating waveform (*

Figure 6.24). The post-processing algorithm also extracted the maximum and minimum resonant frequencies over this period and used them to estimate the change in resonant frequency between each adjacent gas and liquid segment. The results are shown in Figure 6.25. Evidently, there is a fair amount of scatter in this measurement. This is many times greater than the static deviation of resonant frequency estimates, which was less than 10 KHz peak-to-peak. It also occurs almost entirely in the air segment part of the waveform. This scatter is therefore most probably caused by inconsistencies in the air bubble itself, *e.g.,* variations in the thickness of a thin film of toluene remaining on the inside of the capillary during air bubble passage. This retention of a liquid film is to be expected due to imperfections in the contact of the gas segment with the capillary wall, and does not detract from the measurement capability significantly.

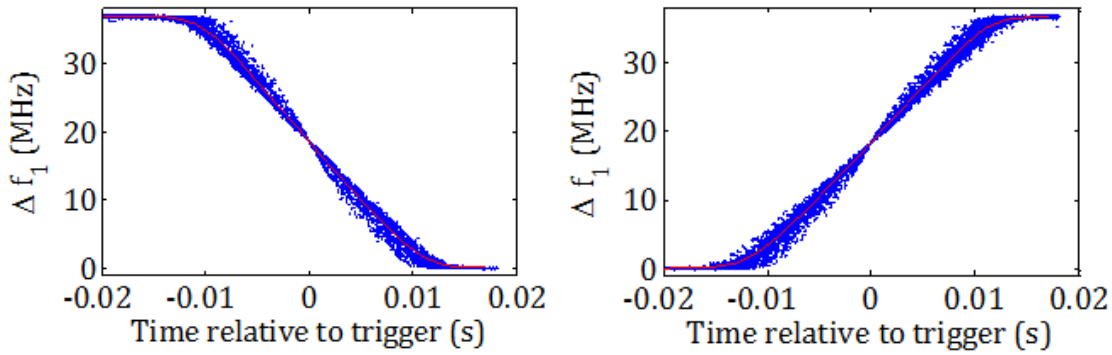



*Figure 6.24: Captured rising and falling edges of the passing of the gas segment showing dispersion caused by variable velocity; the red line shows the average over these 360 segments, blue points are the individual measurements*

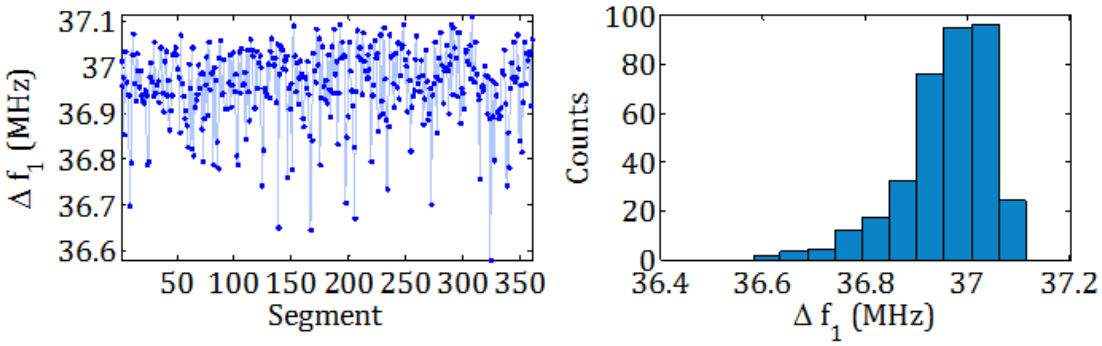



*Figure 6.25: Extracted liquid segment resonant frequency perturbation against time (left column) and histogram (right column)*

## 6.3 Transient microwave heating and control

As discussed in the Applications chapter, the field of microreactor technology is appealing for a number of reasons, not least the promise of high efficiency, high yield on-demand reactions. Many chemical processes require heating, and to ensure optimum yield, purity, *etc.*, heating generally needs to be controlled accurately. Of course, if the only aim is to heat a microreactor, the best and simplest solution is simply to place it in an oil bath. However, part of the promise of microreactor technology is the ability to perform multi-stage reactions and processes on the same 'chip'; each stage may require different thermal environments and may only take up a small area of the chip itself – hence the concept of localised, on-chip heating was born. Whilst there are already excellent general purpose, localised heating methods for microreactors, such as Joule heating via embedded wires, co-flowing channels containing high thermal capacity liquids [4] or micro Peltier effect junctions [5], microwave heating has some specific advantages over other heating methods.

### 6.3.1 Brief review of microwave heating

Microwave heating is a natural partner for so-called 'green chemistry', having demonstrated enhanced yield, higher speed, and better energy efficiency over traditional heating methods in a wide range of processes. This implies reduced waste and reagent consumption and reducing the need for environmentally harmful solvents and reagents. Microwave heating has even been called the "Bunsen burner of the 21st century" [27], and has become the mainstay of organic synthesis [28]. Its unique benefit is selectivity. In a microwave frequency electric field, only substances with high dielectric loss will heat appreciably. This is efficient, as the walls and containment vessel do not waste energy getting hot as well. It can also be exploited to heat a specific phase of a mixture. In a multiphase system, this allows a temperature gradient to be maintained between phases, enhancing certain separation and extraction processes.

Microwaves also penetrate relatively far into even strongly absorbing dielectrics, and they do so instantly from the moment they are applied. This makes microwave heating very fast, as most other externally applied heat sources, such as infrared, only make the surface hot and must wait for heat to diffuse into the full volume of the material. On the micro scale, where convection is negligible, heat diffusion can be glacially slow compared to the timescales of other processes. This fast rate of heating can be further exploited to superheat liquids beyond their normal boiling point – heating simply occurs too fast for the process of bubble nucleation to begin and hence catalyse the liquid-gas phase change. Superheated fluids can be highly reactive and have unusual properties that make them useful for chemical extraction and as reagents [29].


*Figure 6.26: Examples of high-tech single-mode resonant cavity microwave reactors aimed at laboratory synthesis – the CEM Voyager continuous flow system (left) [30] and the Anton Paar Monowave (right) [31] – both offering fibre optic thermometry-based heating control*

Microwave heating apparatus aimed at laboratory work, and even flow chemistry, is already a commercial reality (Figure 6.26), however the combination of directed (*i.e.,* localised) microwave heating with microreactor technology is still at the development stage, with several groups having demonstrated localised microwave heating in microfluidic devices with varying degrees of success. Some of these previous attempts have used broadband transmission line based microwave excitation [32],[33]. This has the advantage of allowing the excitation frequency to

be tuned to the optimum absorbency of specific polar liquids, and has also been used to create specific thermal gradients through standing wave patterns [34]. However, it is inefficient. Energy is dissipated in the on-chip waveguide structures and reflected from the impedance discontinuity created by the presence of the sample in the waveguide. Furthermore, broadband RF signal sources are expensive, so having this broadband capability is not really of practical significance outside of laboratory exploration. Furthermore, materials still absorb microwaves strongly over more than a decade of bandwidth around their relaxation frequency. Other attempts have shown very good control and localisation but have sacrificed the very selectivity of microwave heating that is its main advantage [35],[36].

However, if these issues were solved, applications are not lacking. Microwave heating has been successfully applied to a wide range of lab-on-a-chip applications. These include synthesis processes such as microreactors, [37],[38],[39] drug discovery, [40] microwave-assisted extraction, [41] and bead synthesis, e.g. for molecularly imprinted polymer manufacture [42]. Analytical processes can also benefit from directed microwave heating, such as in microwave-accelerated metal-enhanced fluorescence bioassays, [43] and DNA amplification [44]. The appeal of a well-designed, precision localised microwave heating system is apparent.

### 6.3.2 The self-monitoring microwave project

The mass of fluids contained within a microreactor is generally very small. Heating (and cooling) therefore take place very quickly. Accurate control of heating rate and temperature can therefore be challenging, particularly if it is untenable or undesirable to place thermometric wires, *etc.,* directly at the site where heating is to take place. Any time lag between heating and thermometry will limit the maximum heating rate that the control system can reliably cope with. In a fast-flowing, high-yield microreactor the fluid may be long gone in this time. Whilst fibre optic infrared thermometry is both fast and localised, it is also expensive and is not suitable for large numbers of autonomous parallel reactors or cheap, field-portable instruments; the very implementations of microreactor technology where it can be of most benefit. These considerations lead to the idea of a self-monitoring,

resonant microwave heater. If, by monitoring its resonant frequency and bandwidth, a miniature microwave resonator could regulate and control its own heating, it would mitigate some of these difficulties found in microreactor heating integration. A fast, low-cost, and accurate MRS system such as that strived for in this thesis work could therefore enable the advantages of controlled, localised microwave heating to be integrated with microreactor technology in a cost-effective and practicable way for the first time.

To see if the multitone measurement system could be applied to this heating control problem, a joint research project at Cardiff University was initiated. Its aim was to control the heating of various solvents in a microfluidic chip using a basic microwave cavity resonator, whilst simultaneously measuring the cavity characteristics using the multitone readout system. The project findings were presented at the leading international conference on microfluidic technology [45]. The project went on to investigate ways to adaptively control the optimal coupling of the high power microwave heating signal into the resonant cavity, for which a fuzzy-logic control algorithm was designed. Other members of the project also designed a solid-state amplifier to generate the high power heating signal in a cost-effective and compact way using a laterally diffused metal-oxide-semiconductor (LDMOS) transistor. However, these developments are not relevant to the main discussion of this thesis and will not be covered further.

### 6.3.3 Experiment theory

Dielectric heating in fluids is caused by Joule heating through the movement of ionic charge carriers and the viscous damping of dipole movement induced by a time-varying electric field. The time-averaged heating power $\langle P \rangle$ within a sample of volume $V_S$ and relative permittivity, $\varepsilon_s$, generated by a uniform internal electric field $E_i$ is given by:

$$\langle P \rangle = \frac{V_s \omega_1 \varepsilon_0}{2} \cdot \Im\{\varepsilon_s\} \cdot |E_i|^2. \qquad 6.18$$

Heating is therefore governed by $\Im\{\varepsilon_s\}$, the sample's dielectric loss factor. The fact this is material-specific means that selective and efficient heating is possible. Assuming negligible heat loss (valid for rapid heating in a thermally insulating channel), the heating rate of a material is therefore:

$$\frac{dT}{dt} = \frac{\langle P \rangle}{\rho C} = \frac{V_s \omega_1 \varepsilon_0}{2\rho C} \cdot \Im\{\varepsilon_s\} \cdot |E_i|^2, \qquad 6.19$$

where $\rho$ is the sample density and $C$ its specific heat capacity [46].

The problem with using this simple analysis to estimate and control the temperature accurately is the difficulty in estimating $|E_i|$. This factor depends on the specific resonant structure and its mode, as well as sample position, sample shape, and the relationship between the power delivered from a source and the electric field amplitude within the cavity. For example, if the applied electric field, $E_a$, crosses a boundary between regions of different permittivity, depolarisation will cause the internal electric field, $E_i$, to be reduced. Just as this shielding effect causes the change in bandwidth to be much less under depolarisation conditions than without, so a depolarised sample will heat much less efficiently. Trying to heat a fluid in a round capillary (if the split ring resonator were to be used as a miniature microwave, for example) the heating power would be reduced by the factor

$$\frac{|E_i|^2}{|E_a|^2} = \frac{|4\varepsilon_1|^2}{|(\varepsilon_1 + \varepsilon_2)(\varepsilon_1 + 1) - (\varepsilon_2 - \varepsilon_1)(\varepsilon_1 - 1)(a/b)^2|^2}, \qquad 6.20$$

assuming the same power is coupled into the resonator. The same would apply to heating droplets, as was attempted by [33]. In this case, the heating power density within the water droplets (with $\varepsilon_s \approx 80 - j10$ at 2.5 GHz) in the low loss fluorocarbon oil used ($\varepsilon_m \approx 2$), would have been about 0.5 % of that possible with a continuous water column in a non-depolarising geometry, *i.e.,* where the electric field is applied parallel to the channel axis. Not only would this require a much higher power to deliver the same heating rate, it would also cause a larger portion of incident microwave power to be wasted in the rest of the resonant structure,

making it less energy efficient. The difference between attempting to heat microfluidic channels in non-depolarising as opposed to a depolarising regime is easily demonstrated by a simple FEM simulation (Figure 6.27).

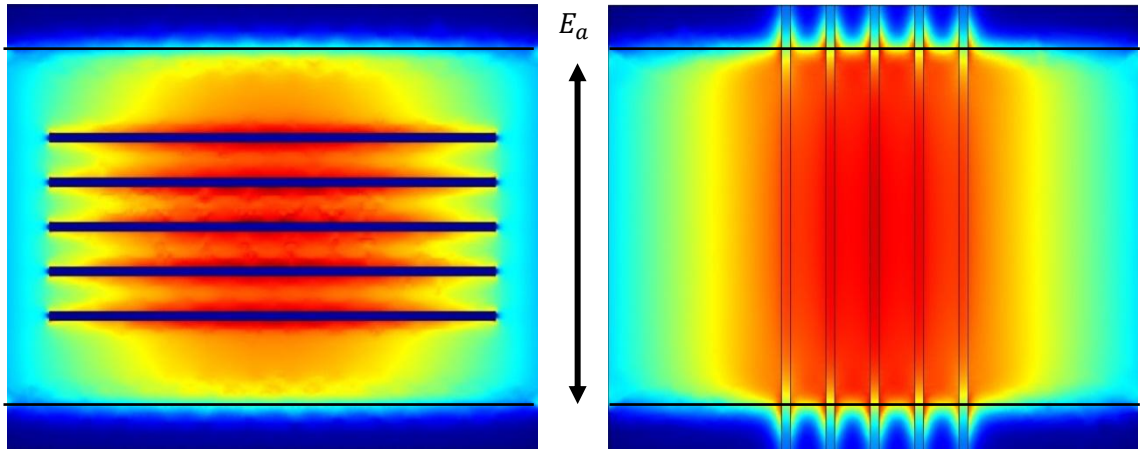



*Figure 6.27: Section through five parallel rectangular channels of water, perpendicular (left) and parallel (right) to the direction of applied electric field in a microwave cavity resonator, with electric field strength shown as a colour gradient from dark blue (zero) to red (maximum). The black line at the top and bottom of each figure marks the limit of the cavity and the beginning of the slot in the cavity wall – an evanescent field region.*

In this experiment, the channels where kept vertically aligned with the direction of the electric field to minimise depolarisation. The perturbation response of this rectangular cavity containing $L$ vertical channels is therefore:

$$\frac{p(\theta)-p_0}{p_0}=\frac{-\sum_{k=1}^{L}V_k|E_k|^2\varepsilon_s}{2V_e|E_0|^2+\sum_{k=1}^{L}V_k|E_k|^2\varepsilon_s}, \qquad 6.21$$

where $E_k$ is the electric field amplitude at the position of the channel, assumed uniform over its length. This assumption is invalidated close to the top and bottom of the cavity where the channels enter and leave it via a slot – as shown by the FEM simulation above. The deviation from uniform field depends on the permittivity. For these simulated channels containing water, the effect of depolarisation at the extremities of the cavity can be seen around the evanescent field in slot region, reducing the electric field in the channels relative to that around them. This zone of

sample-dependent field will invalidate the static field assumption and hence render a linear perturbation model inaccurate, as discussed in the Theory chapter.

Reducing the width of the slot, through using a thinner microfluidic chip, would reduce the relative significance of this effect. However, despite attempts to fabricate PTFE microfluidic devices from thinner PTFE sheets, (using pre copper clad Polyflon® substrates [47] stripped of copper) the bonding process was not as reliable, and the micromilling used to cut out the channel profiles was poorly controlled due to the difficulty in keeping this thin substrate flat over the course of the milling process. However, it was found recently that laser cutting of the bonding film allows for tighter control of the bonding process [48], so further work might improve this method sufficiently, eventually allowing thin, flexible pure PTFE microfluidic devices to be manufactured as desired.

For the $TE_{101}$ mode, the electric field distribution in the Cartesian coordinates *x, y, z,* is very simple, hence the effective volume can be deduced:

$$V_e E_0^2 = \int |E(0)|^2 \, dV = E_0^2 \int_0^a \int_0^b \int_0^d \sin\frac{\pi x}{a} \sin\frac{\pi z}{d} \, dx \, dy \, dz = \frac{E_0^2 V_c}{4}, \qquad 1.$$

where $V_c = abd$ is the cuboid volume of the cavity. Thus for a single channel or capillary placed in the position of highest electric field, narrow enough to be able to assume the field is uniformly $E_0^2$ over its entire length, the perturbation equation reduces to the very simple form

$$\frac{p(\theta) - p_0}{p_0} = \frac{-2V_s\varepsilon_s}{V_c + 2V_s\varepsilon_s}. \qquad 6.22$$

### 6.3.4 Linking resonant sensing to microwave heating

The amount of heating power delivered to a sample can be estimated in a resonator by considering the basic definition of the imaginary part of the resonator eigenvalue, $\sigma_0$, (with the sign taken to be negative by convention) which can be expanded into a summation over *N* regions due to the additive nature of energy. In

principle, this allows the power dissipated in a specific region, $\langle P_j \rangle$, to be found by measuring the unloaded bandwidth:

$$\sigma_0 \triangleq \frac{\langle P \rangle}{2\langle U \rangle} \quad \rightarrow \quad \sigma_0 = \frac{\sum_{i=1}^{N} \langle P_i \rangle}{2\sum_{i=1}^{N} \langle U_i \rangle} \quad \Rightarrow \quad \langle P_j \rangle = 2\sigma_0 \sum_{i=1}^{N} \langle U_i \rangle \quad - \sum_{i=1,\ i \neq j}^{N} \langle P_i \rangle \qquad 6.23$$

However, in practice the stored energy $\langle U \rangle$ is unknown. A simple solution is to use the simple 'zeroth order' perturbation assumption that the stored energy remains the same when perturbing factor, $\theta$, is introduced into the resonator. Then the ratio of the power dissipated before and after the perturbation can be related to the bandwidth:

$$\frac{\langle P(\theta) \rangle}{\langle P(0) \rangle} = \frac{\sigma_0(\theta)\langle U(\theta) \rangle}{\sigma_0(0)\langle U(0) \rangle} \approx \frac{\sigma_0(\theta)}{\sigma_0(0)}. \qquad 6.24$$

Furthermore, if the perturbation is actually the act of introducing a sample into the resonator, then the above result can be used to derive the ratio of the power dissipated in the sample, $\langle P(\theta) \rangle - \langle P(0) \rangle$, to the total power dissipated throughout the resonator. In other words, this is the power efficiency, $\eta$, of sample-specific heating within the resonator:

$$\eta_{int} = \frac{\langle P(\theta) \rangle - \langle P(0) \rangle}{\langle P(\theta) \rangle} \approx 1 - \frac{\sigma_0(0)}{\sigma_0(\theta)}. \qquad 6.25$$

This gives the relative power dissipated in the sample compared to that dissipated in the metallic cavity walls, microfluidic substrate, *etc*. It also means that the relative power dissipated in a sample can be estimated just by measuring the bandwidth, as long as the bandwidth before the sample is introduced is known. Also, if the heating power is assumed to be coupled into the resonator at a single port, then the total power dissipated within the cavity, $\langle P(\theta) \rangle$, can be found by measuring the incident and reflected power at that port, *i.e.,* $\langle P(\theta) \rangle = \langle P_{in} \rangle (1 - |\Gamma|^2)$, where $P_{in}$ is the power produced by the amplifier/magnetron source and $\Gamma$ is the reflection coefficient. This is a parameter that can be measured in realtime,

using a VNA or a directional coupler and diode power detector, giving a measure of the instantaneous 'external' power efficiency of heating:

$$\eta_{ext}(t) = \frac{\langle P(\theta)\rangle}{\langle P_{in}(t)\rangle} \approx \left(1 - \frac{\sigma_0(0)}{\sigma_0(t)}\right)(1 - |\Gamma(t)|^2). \qquad 6.26$$

From the Theory section analysis, the reflection parameter at the resonant frequency (the optimum frequency for heating excitation) is:

$$S_{11}(\omega) = K_0 + \frac{K_1 Q_1}{1 + j\frac{2Q_1}{\omega_1}(\omega - \omega_1)} \Rightarrow \ |\Gamma|^2 = |K_0 + K_1 Q_1|^2 \ \ (\omega = \omega_1). \qquad 6.27$$

Using the small coupling reactance assumptions, and assuming the resonator is one-port with magnetic (loop) coupling, with $K_0 \approx -1$, gives:

$$|\Gamma|^2 \approx \left|-1 + \frac{2Q_1}{Q_e}\right|^2 = \frac{(2Q_1 - Q_e)^2}{Q_e^2} = \left(\frac{\omega_1 x_c k_m^2}{\sigma_1} - 1\right)^2 \qquad 6.28$$

This shows that the reflection parameter is dependent on the sample, as varying sample permittivity will cause the $Q$ factor to change. However, it also shows that in principle the external efficiency could be measured without using a reflectometer as long as $Q_e$ was fixed and its value previously characterised. From the above equation, it is apparent that the highest heating efficiency is obtained when $2Q_1 = Q_e$, or $Q_0 = Q_e$. This is the 'critical coupling' criterion. In the experimental cavity, the coupling was designed to be altered by rotating a circular coupling loop; allowing it to be optimised for maximum efficiency. This allows the coupling to be varied as the mutual inductance between the cavity and the loop $M_{12} \propto \int H \cdot dS = H_0 \cos\theta$, where $\theta$ is the angle between the (horizontal) magnetic field lines at the site of the loop and the loop axis. Thus $k_m^2 \propto \cos^2\theta$ and the efficiency can be written

$$\eta_{ext}(t,\theta) \approx 1 - \left(\frac{\omega_1(t)}{\sigma_1(t)} K_1(0) \cdot \cos^2\theta - 1\right)^2, \qquad 6.29$$

giving the optimum angle for maximum power transfer at any given time,

$$\hat{\theta}_o(t) = \cos^{-1}\sqrt{\sigma_1(t)/\omega_1(t)K_1(0)}, \qquad 6.30$$

although the constant $K_1(0)$ still needs to be found by prior characterisation for this equation to be accurate. An alternative approach is to use an iterative method to minimise the reflected power, although this can be a tricky nonlinear problem when coupled with the task of trying to keep the frequency of power excitation on resonance as well. This issue is outside of the scope of this discussion, however.

For uniform, rapid heating of a stationary fluid, where the transport of heat away by convection and conduction can be neglected, this suggests that the temperature rise could be predicted from resonator bandwidth measurements:

$$T(t) - T(0) \approx \frac{1}{\rho C}\int_0^t \left(1 - \frac{\sigma_0(0)}{\sigma_0(t)}\right) \eta_{ext}(t)\,\langle P_{in}(t)\rangle \cdot dt. \qquad 6.31$$

This time integral can be approximated by the cumulative sum of successive discreet samples, as long as the rate of change is slow relative to the sampling rate. If the liquid is flowing, this will remove heat at a constant rate from the system. This situation can also be accounted under quasi-steady-state conditions, where $\sigma_0(t)$ and $\eta_{ext}(t)$ are assumed to be changing slowly, by evaluating the above integral to approximate the temperature rise at the outlet of a length of channel:

$$T_{out} - T_{in} \approx \frac{\ell}{v\rho C}\left(1 - \frac{\sigma_0(0)}{\sigma_0(t)}\right) \eta_{ext}(t)\,\langle P_{in}(t)\rangle, \qquad 6.32$$

where $v$ is the volumetric flow rate and $\ell$ is the channel/capillary length in the uniform electric field. This implies a linear temperature gradient along the channel length. However, this steady state solution is invalidated for larger temperature changes by the temperature dependence of the complex permittivity of the liquid itself. In the above equations, the power dissipated in the sample is temperature dependent, $\langle P(T)\rangle \propto \Im\{\varepsilon_s(T)\}$, making the heating problem nonlinear and creating a non-uniform rate of heating with distance along the channel/capillary . Generally,

most polar liquids show a permittivity-temperature relationship that is dominated by the reduction in the dielectric relaxation time with temperature; the higher thermal energy of the molecules within it allowing the displacement caused by the microwave electric field to be damped away more quickly. This activation process, where the molecular relaxation mechanism is characterised by a certain activation energy $E_a$ [49]

$$\tau = \frac{\tau_0}{T} \exp\left(-\frac{E_a}{k_B T}\right). \qquad 6.33$$

The frequency corresponding to maximum power absorption will therefore shift upwards with temperature. Thus, if the absorption maximum is higher than the excitation frequency to begin with, both the real and imaginary permittivity will drop. This is known as 'self-limiting' thermal behaviour because the rate of heating decreases as the material heats up. However, if the frequency of excitation is initially higher than the absorption maximum, the imaginary permittivity can increase with temperature, leading to 'thermal runaway' behaviour as the temperature increases exponentially with time. Fortunately, the temperature dependence of permittivity is known for a number of common solvents, especially water, where it is of interest in the study of its intermolecular structure [50],[51].

### 6.3.5 Experiment setup

The heating cavity/MRS was a simple $TE_{101}$ mode rectangular resonator, fabricated by milling from an aluminium block and joined vertically at the midpoint to avoid current flow across this join. The rectangular $TE_{101}$ mode is functionally equivalent to the $TM_{010}$ mode in a cylindrical resonator, having a uniform electric field along the vertical axis where the sample is placed with good magnetic field separation. A slot was milled to allow insertion of a 90 mm × 60 mm × 4 mm PTFE microfluidic 'cartridge'. The cavity had three coupling ports. One large, rotatable magnetic loop coupling allowed critical coupling of the heating signal into the cavity. Two secondary SMA capacitive probe coupling ports allowed connection to the MRS system measuring the cavity in transmission. This design decoupled any variation

in the high power coupling from the MRS system and provided good attenuation of the high power signal from the measurement system. However, it also introduced additional sources of coupling loss and may have been somewhat redundant – ideally, both heating and multitone excitation signal would be coupled in via the single adjustable port. Further details of the resonator design and of the microfluidic chips and their manufacture are discussed in [44].

Microwave power up to 2 W was provided by a synthesiser (NI PXI-5652) combined with a 35 dB solid state power amplifier (PA) and was coupled into the cavity resonator using the adjustable magnetic coupling loop. The MRS readout system used was the first implementation of the multitone system (refer to the Hardware chapter for details). Because of the limited characterisation bandwidth of this system, the microfluidic chip caused a significant increase in the loaded bandwidth, making the resonance difficult to characterise. Thus in the experiments discussed here a single vertical Teflon capillary, 1/16" OD and 500 μm ID (Sigma Aldrich, UK), located along the central axis of the resonator was used rather than the microfluidic chip. After all, the aim as regards this thesis is to demonstrate the use of the multitone system, rather than to show explicitly that the method can be applied to an actual microfluidic device. Whilst the second implementation of the multitone system would have been more than adequate, it could not be used for this demonstration due to time constraints.

Control of heating was achieved by on-off modulating the continuous-wave heating signal using a PIN diode switch controlled by a variable duty cycle signal (internal to the RF signal generator used as the source). This pulse width modulation (PWM) was at sufficiently fast rate (20 Hz) that the time averaged source power, $\langle P_{in}(t) \rangle$, can be assumed to be proportional to the duty cycle, and, due to the thermal time constant of heating, should result in almost uniform heating rate with no significant transient. This method of heating regulation is the most energy efficient, as the PA drive level can be optimised for best efficiency. In a final product, it is possible that a switched-mode power amplifier could be used instead for even greater power efficiency. It is also accurate, as the incident power can be estimated

accurately without explicit measurement once the gain of the PA is characterised. This saves additional circuit complexity and signal processing.

In order to test the hypothesis that temperature regulation should be possible using just the MRS readout system without any secondary thermometry, a simple control loop was implemented in software. This control loop implemented basic proportional control and was not optimised. The control signal could be either resonant frequency or the bandwidth. Target resonant frequency or bandwidth values (*i.e.,* thermostatic) or target waveforms (*i.e.,* thermal cycles) for these variables could then be input into the algorithm, which would then adjust the duty cycle of the heating signal modulation and hence regulate the rate of heating.

For these control experiments HPLC grade water (Sigma Aldrich, UK) was used. Water heats very efficiently at 2.45 GHz, but quite slowly as it has a high heat capacity. It also shows a significant temperature dependence of the real part of permittivity, due to the relaxation peak (initially at 18 GHz) moving to higher frequencies as it heats up. Resonant frequency control was therefore used, based on a locally linear approximation for the permittivity dependence of water and the knowledge that at the excitation frequency of 2.45 GHz, the slope of the permittivity-temperature relationship is always negative. Thus, if the measured resonant frequency increased, it meant the permittivity was decreasing and hence the temperature was rising. The heating duty cycle was thus accordingly reduced. The constant of proportionality used was merely adjusted empirically to give the desired temperature range needed. However, for aqueous samples it is entirely possible that, with prior calibration using a reference thermometry standard such as infrared emissivity via fibre optics, relatively accurate temperature control could be possible. As far as I know, this is the first time such a heating control method has been demonstrated.

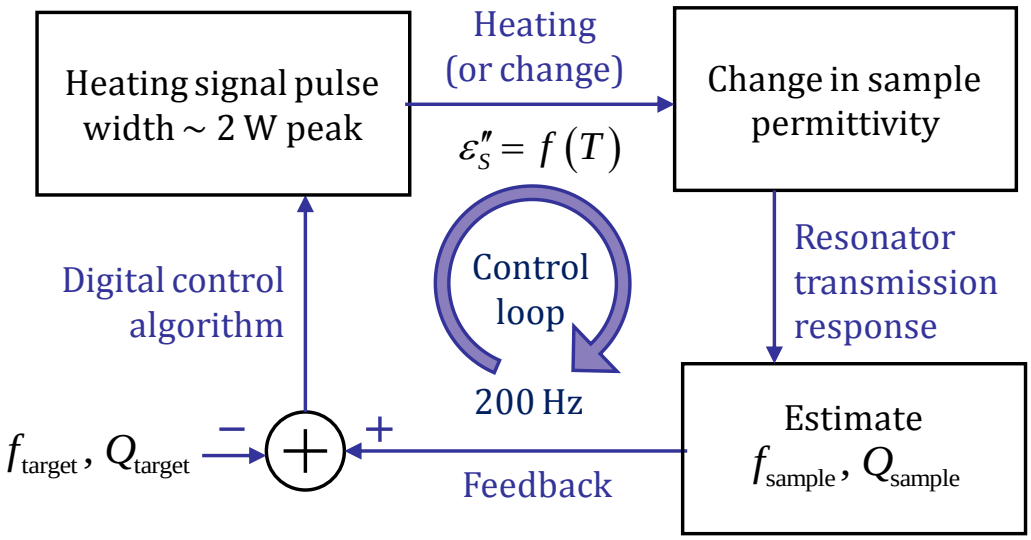



*Figure 6.28: Microwave heating control cycle using resonant frequency and bandwidth feedback from the MRS readout system*

The system architecture is shown in a simplified form below, although in reality there were three separate sources all nominally tuned to the same frequency, rather than the single RF source shown here. This was due to the restrictions of instruments used to implement the initial multitone system. Furthermore, an additional isolator and attenuators were required at the output and input of the multitone system to protect it from the high power heating signal that is still of a significant amplitude when coupled into the secondary ports.

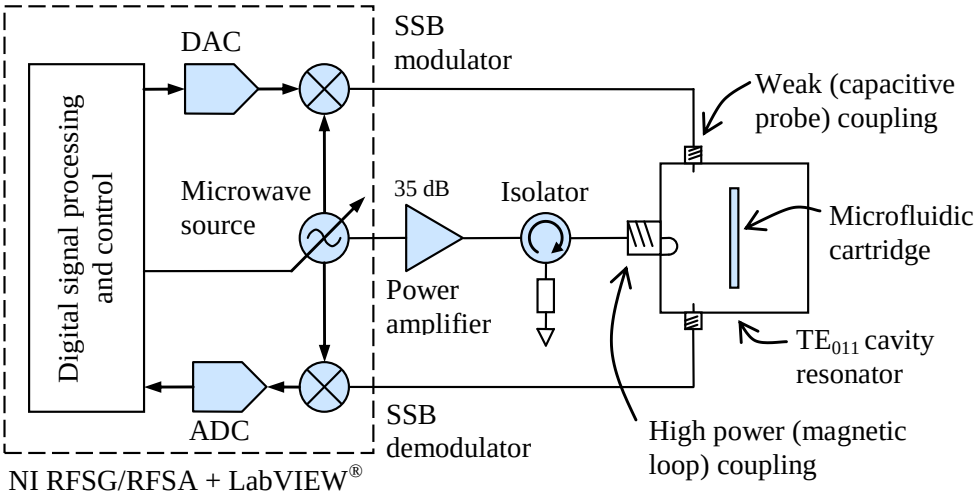



*Figure 6.29: Architecture of the microwave heating system with simultaneous MRS readout based on the multitone system*

In order to prevent the high power heating signal from swamping the measurement system, readings were only taken during the 'off' phase of the PWM cycle. This meant that a duty cycle above about 80 % was not possible due to the time needed to acquire the measurement waveform. However, this only influences the maximum heating rate and is not a significant disadvantage. One issue with the restrictive system implementation was that synchronisation of the ADC acquisitions with the PWM signal had to be done in software. This made the sampling timing inaccurate and resulted in the overall measurement rate being limited to a disappointing 50 ms. Future iterations of this heating control system based on the second implementation of multitone system would alleviate this, as well as many other limitations of this setup.

### 6.3.6 Results of dynamic heating experiments

As an initial test, a single full-power (2 W) pulse was applied to water under a stopped flow condition, and the resonator response was measured using the multitone system. The simple single capillary perturbation model was used to invert these data to complex permittivity. Despite its innacuracies this model was sufficient for an indicative demonstration. The results (Figure 6.30) show that the

change in complex permittivity as the temperature rises can be measured in realtime with the system simultaneously to the application of microwave heating.

The large amount of noise on these measurements is disappointing, and may be due to noise and interference amplified by the PA being coupled into the receiver. The noise is certainly greater than that obtained with the nominally static measurements discussed in the previous two chapters. One issue is apparent in these results. 'Parasitic' heating effects (due to the change in the permittivity of the capillary, thermal expansion of the cavity *etc*.) cause systematic error. This can be seen in the difference in the apparent sample permittivity before and after the heating pulse is applied. This error is approximately the same magnitude as the unexpected change in permittivity that occurs towards the end of the heating cycle. What this implies is that, although the water in the capillary can be seen to heat up and cool down relatively quickly, (albeit with the expected decrease in heating rate due to its lower loss factor with temperature) it does not return to the same permittivity even by the end of the experiment.

Given that the permittivity of the capillary is low, that Teflon shows relatively little temperature dependence, and that the apparent time lag is consistent with a long thermal time constant, this parasitic perturbation is probably caused by the air heating up inside the cavity and causing the water to stay hot. The capillary thermal mass would be too low to cause this alone. The simple solution to this problem, therefore, is to provide a constant flow rate of cooling air, or continuously flow another microwave transparent fluid, such as perfluorohexane, along a coaxial tube or parallel channel. This would provide more rapid cooling for fast thermal cycles, at the expense of heating rate, and would result in a more accurate heating rate estimation from bandwidth measurements, as the rate of heat removal could be accounted for and carefully controlled.

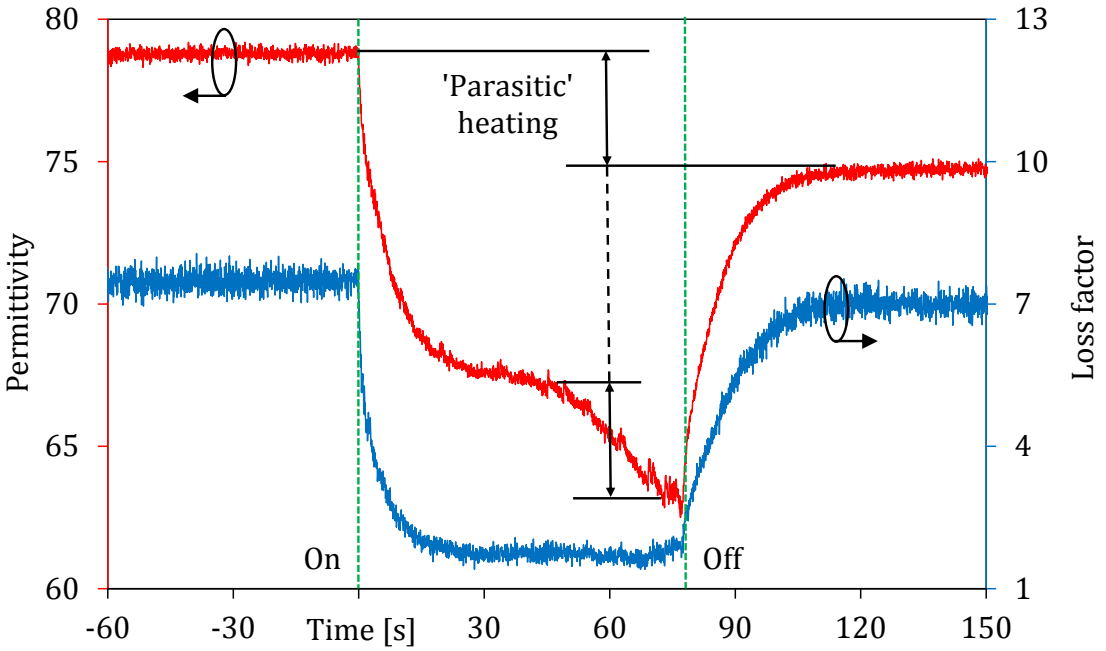



*Figure 6.30: Measured permittivity response of a capillary containing water in a stopped-flow condition subject to a full power heating pulse.*

In the next two results, the control system was used to create arbitrary thermal cycles. In the first, a sinusoidal 'target' wave in resonant frequency was input into the control algorithm. The water in these experiments was flowed at high constant rate of 0.5 ml/min using a syringe pump. It can be seen that the temperature measured by a thermocouple at the outlet of the cavity does appear to follow a sinusoidal pattern, lagging behind the measured resonant frequency by a small amount due to the thermal time constant of the thermocouple. In the second example a larger signal control waveform was used to create a dynamic temperature swing of 20°C. In both of these experiments, an expectedly high level of noise is present on the resulting resonant frequency measurements. This is in contrast to the previous heating pulse experiment, which in all other respects apart from the control loop was identical. Therefore, the source of this additional noise must be the control loop itself. This is probably caused by the simplistic control algorithm, which amplifies the natural resonant frequency fluctuation. Implementing a proper control algorithm with empirically optimised proportional, integral, derivative (PID) settings would alleviate this issue.

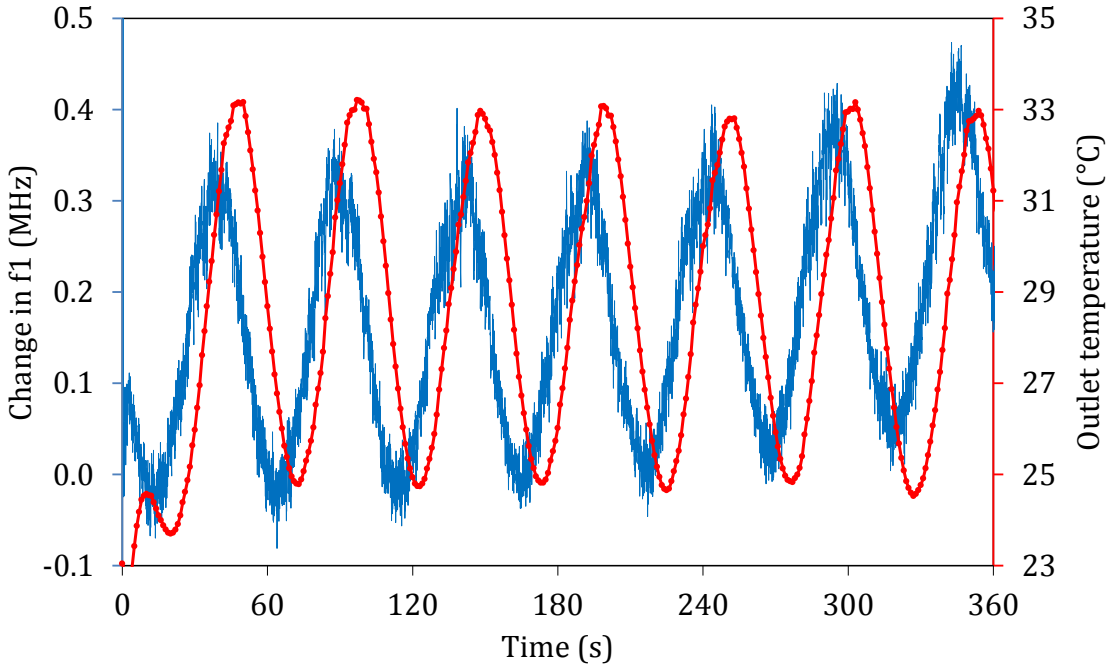



*Figure 6.31: Heating cycle where a sinusoidal target variation in resonant frequency is used as a heating control signal. Red shows the temperature measured at the outlet, blue the measured resonant frequency.*

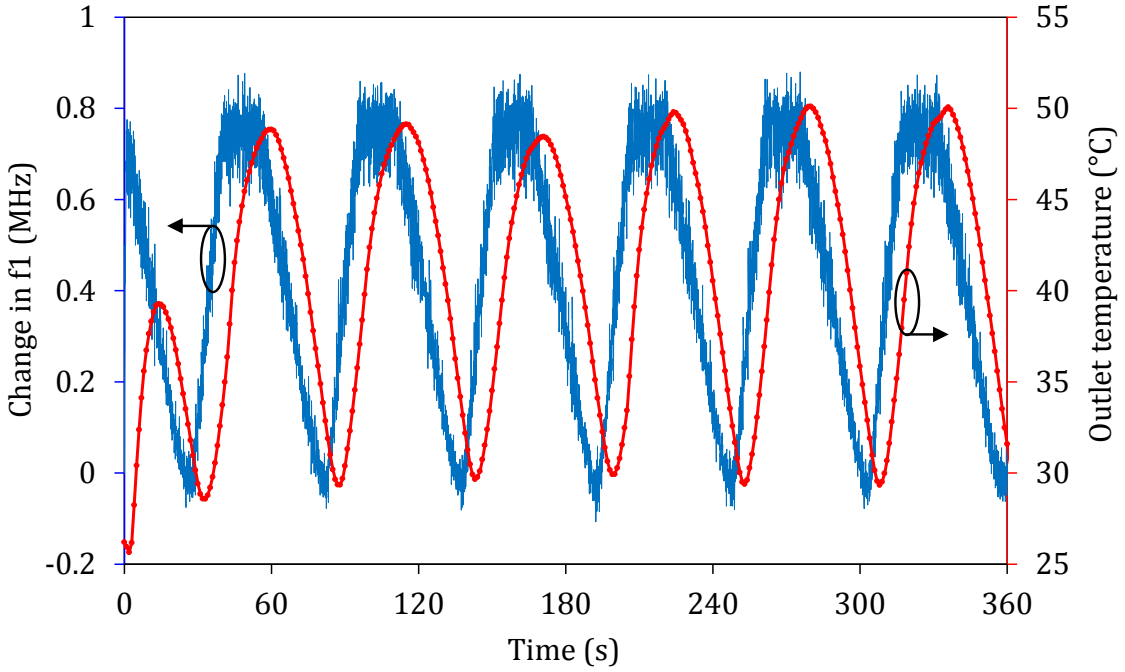



*Figure 6.32: Higher amplitude heating cycle, where a ramp -hold waveform in resonant frequency is used as a heating control signal. Red shows the temperature measured at the outlet, blue the measured resonant frequency.*

## 6.4 Dynamic capillary filling

The idea to measure the time-domain filling of a capillary using a microwave resonator arose out of on-going work at Cardiff University investigating blood chemistry using microwave dielectric analysis. To support this work a joint project with visiting students from L'Université de Bordeaux was undertaken to investigate the metrology of capillary filling using dynamic microwave resonant sensor measurement [51]. This project represents a good example of the power of temporal MRS measurement to extract more information about a sample than would otherwise be possible. Due to the requirements for accurate time-sampling and the relatively slow rate of filling, the optimised VNA readout discussed in the Hardware chapter was used for these experiments.

Capillary affinity is the basic principle on which most chromatography techniques are based. For example, in thin layer chromatography (TLC) an analyte is drawn up by capillary action and is separated into different components by the distance they travel in a planar substrate, governed by their affinity for the substrate material. Whist provoking comparison to simple school experiments with ink and blotting paper, modern TLC is highly sophisticated and is regarded as a fast, inexpensive, flexible, and low-footprint analysis method [52]. Although simple paper based immunoassays (most commonly seen as pregnancy test kits) are low-cost and easy to manufacture, such tests are difficult to apply in a quantitative way [53]. Perhaps it is not too much to suppose that a low-cost MRS system has the accuracy to make such tests quantitatively accurate, without necessitating complex microfluidic fabrication, pumping systems or expensive, fragile detectors.

The concept of applying a detection method distributed spatially along a capillary (chromatography 'column'), rather than at a single point after sample elution, has been applied to pumped-flow chromatography as well, where it is known as whole-column detection [54]. The study of the dynamic capillary filling properties of an analyte has elements of both these approaches. Due to the correlation between distance and time in the filling process, a non-homogenous chromatographic separation will cause the time-domain filling profile to deviate

from that of a pure liquid, potentially providing information about complex samples in an extremely simple way.

Combining chromatographic separation with microwave resonant sensor measurement is unexplored territory, and MRSs could potentially add a new dimension to analyte measurement. When combined with accurate, fast, and low cost MRS readout technology, a chromatography-MRS hybrid system could offer advantages in cost, selectivity, or sensitivity in specific applications over conventional detection methods (such as conductivity, UV, visible or fluorescence spectroscopy). The measurement of capillary filling dynamics with MRSs is therefore worthy of investigation, at least at the proof-of-principle level as done here.

### 6.4.1 Experiment theory

The interface between a solid and a liquid is characterised by the relative magnitude of the forces acting to keep the liquid molecules together (cohesive forces, *e.g.,* van der Waals, hydrogen bonding, *etc.*) and attracting the molecules to the solid surface (adhesive forces). The traditional way of characterising this interface between a specific liquid and specific solid is by considering the contact angle made with a small drop of the liquid resting on a flat surface of the solid – the so-called sessile drop technique [55],[56]. When at rest this three-phase (gas, liquid, solid) system is in thermodynamic equilibrium with all interfacial energies between each phase in balance; a situation described by the Young equation [57]. Liquids with a high affinity for the solid are said to have good wetting; the drop spreads out further and therefore has a low contact angle at the point where the liquid-gas interface meets the solid-liquid interface (Figure 6.33). The study of wettability is an active area of research [58].

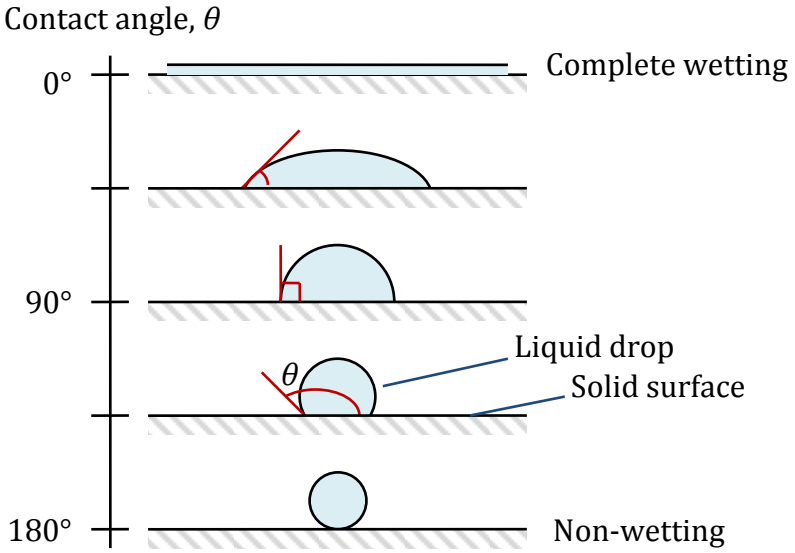



*Figure 6.33: Possible behaviours of a sessile droplet on a surface and the definition of contact angle – after [56].*

If a capillary is made of a materials that has a contact angle less than 90° with a particular fluid, this fluid will be drawn up inside it by adhesive forces – the phenomenon known as capillary action. The filling process (capillary imbibition) itself is an interesting dynamic phenomenon which is still an important area of research of importance to subjects such as botany and microfluidics. Capillary forces attract the liquid up the tube but viscous forces, gravity, and inertia all impede its motion. The dynamics of a vertical capillary filling a distance $x$ in time $t$ with a Newtonian liquid under a laminar flow regime (Poiseuille flow) are described by the differential equation [59],[60].

$$\underbrace{\rho x\frac{d^2x}{dt^2}+\rho\left(\frac{dx}{dt}\right)^2}_{inertial}+\underbrace{\frac{8\eta}{r^2}x\frac{dx}{dt}}_{viscous}+\underbrace{\rho gx}_{weight}=\underbrace{\frac{2\gamma}{r}\cos\theta,}_{capillary} \qquad 6.34$$

$\rho$ and $\eta$ are the density and viscosity of the liquid, $\gamma$ is the surface tension of the liquid-air interface, $\theta$ is the liquid-capillary contact angle and $r$ is the capillary radius. Exact analytic solutions to this equation are not available. However, by ignoring the inertial terms, capillary forces compete only with viscous forces to draw liquid up the capillary unaided. This solution was independently exposed by

Lucas [61] and Washburn, [62] and is known as the Lucas-Washburn (often just Washburn) equation

$$x^2(t) - x^2(0) = \frac{r\gamma}{\eta} \cdot \frac{\cos\theta}{2} \cdot t. \qquad 6.35$$

This equation predicts a $\sqrt{t}$ dependence for the distance the liquid meniscus moves in a given time. However, this equation does not converge to the steady state height,

$$\lim_{t\to\infty} x(t) = \frac{2\gamma\cos\theta}{\rho g r}, \qquad 6.36$$

and it also predicts a nonphysical infinite velocity at the start of filling. Furthermore, the contact angle itself may not be constant but depends to some extent on the velocity of filling, due to friction effects at the moving meniscus contact line [63]. Despite this, the Washburn equation is simple yet relatively accurate in predicting the dynamic filling behaviour over the majority of filling progression.

If capillary filling is to be measured by resonator perturbation, if the sample is assumed to have a constant permittivity and is non-depolarising as is the case here, a simple approximation is to assume the change in resonant frequency (or bandwidth) is linearly related to the height of the meniscus in the cavity:

$$\frac{f_1(t) - f_1(0)}{f_1(\infty) - f_1(0)} = \frac{x(t) - x(0)}{L + x(0)}, \quad (x(0) < x(t) < L + x(0)) \qquad 6.37$$

where $L$ is the internal height of the cavity and $x(0)$ is the distance between the bottom of the capillary and the point at which the meniscus just enters the resonator cavity. $f_1(\infty)$ is the value of the resonant frequency when the liquid fills the resonant cavity fully, and $f_1(0)$ is the resonant frequency before the sample enters. Thus, given the $\sqrt{t}$ dependence of the displacement, a suitable model for the time dependence of the relative change in resonant frequency is:

$$\frac{f_1(t) - f_1(0)}{f_1(\infty) - f_1(0)} = \frac{\sqrt{Dt} - x(0)}{L + x(0)}, \tag{6.38}$$

where $D \approx r\gamma \cos\theta / 2\eta$ is the diffusivity constant of the liquid filling process. Note that, due to the initial length $x(0)$, the squared change in resonant frequency will not be a straight line, as might be initially be expected, since

$$\left(\frac{\Delta f_1(t)}{\Delta f_1(\infty)}\right)^2 = \frac{Dt + x(0)^2}{\left(L + x(0)\right)^2} - 2\frac{x(0)\sqrt{Dt}}{L + x(0)} = a_1 \cdot t + a_2 \cdot \sqrt{t} + a_3 \tag{6.39}$$

This last equation will prove useful in curve fitting the filling process in a numerically robust manner, giving $D = a_2^2 / 4a_3$.

Studying the capillary filling process in ways other than simply measuring the meniscus position optically can reveal useful secondary information about the liquid and its properties, especially if there are other time-domain factors that perturb the normal filling process. For example, if the capillary is coated with a chemical that reacts with the liquid, or has a certain affinity for different components within the liquid.

### 6.4.2 Experiment setup

To investigate capillary filling metrology with microwave resonant sensors, a simple experiment was devised based around a $TM_{010}$ mode cylindrical resonator designed to accept a rigid, narrow capillary along its axis. The basic experimental setup is shown diagrammatically below:

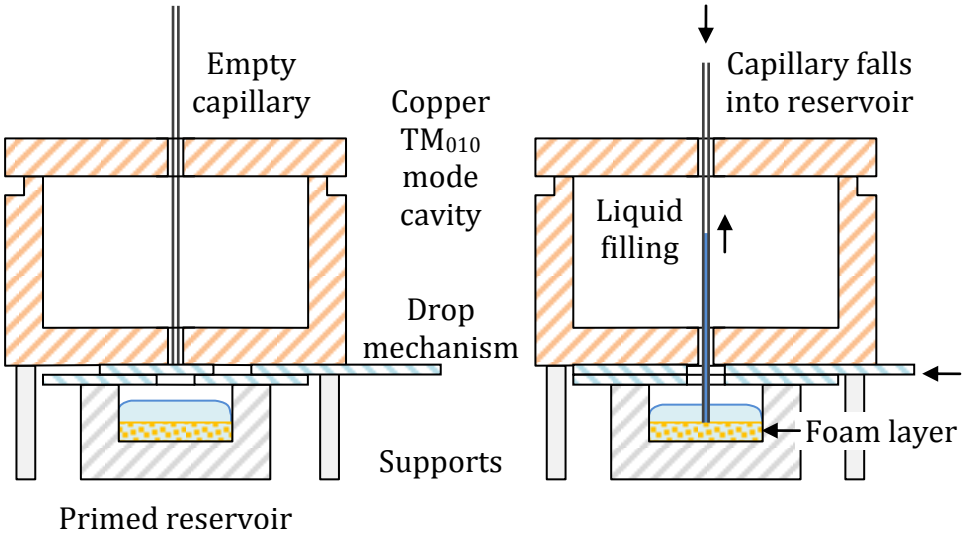



*Figure 6.34: Cross sectional diagram of the transient capillary filling experimental apparatus showing the primed experiment (left) and during the filling transient (right) (not to scale).*

A cylindrical copper $TM_{010}$ mode resonator of internal diameter 33 mm and height 24 mm was designed and fabricated for the project, giving a resonant frequency of around 7 GHz. The resonator was conventionally turned by lathe and was fastened together with through-bolts. This meant that the radial-longitudinal current flow in the $TM_{010}$ mode was forced to cross the join between the lid and the main resonator body. In order to reduce the impact of this on the unloaded Q, a small 1 × 1 mm lip was introduced in order to magnify the contact force and provide a better electrical connection at this point. A 1 mm diameter hole along the central axis permitted various sizes of capillary to be aligned to the uniform electric field maximum. Any homogeneous dielectric-filled capillary would therefore be non-depolarising with this arrangement. However, during the filling process a dielectric discontinuity is created by the rising meniscus, which causes depolarisation at the liquid-air interface, resulting in a degree of nonlinearity in the perturbation vs. distance relationship as a function of permittivity. In order to determine and correct for this effect, parametric FEM simulation was used. Due to the axisymmetric geometry, a 2D solution was applicable, which was very efficient in terms of computational burden and allowed an exceptionally fine mesh to be used, effectively rendering the quantisation error negligible.

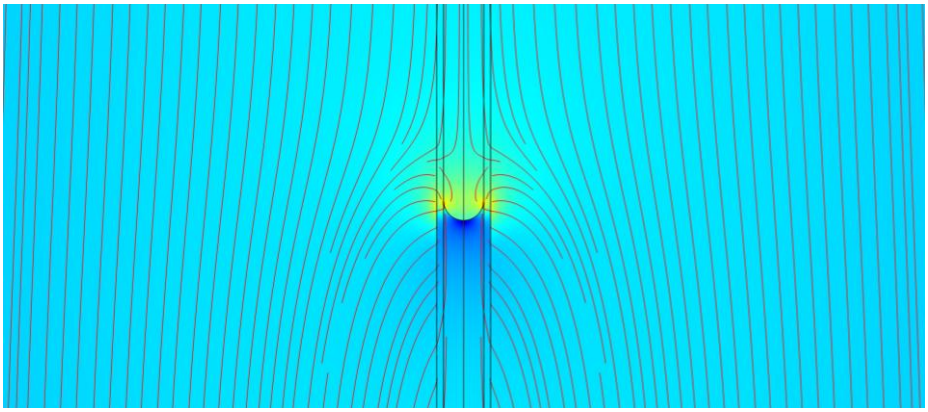


*Figure 6.35: Depolarisation around the moving meniscus, showing |E| shaded from blue-red with red E field lines*

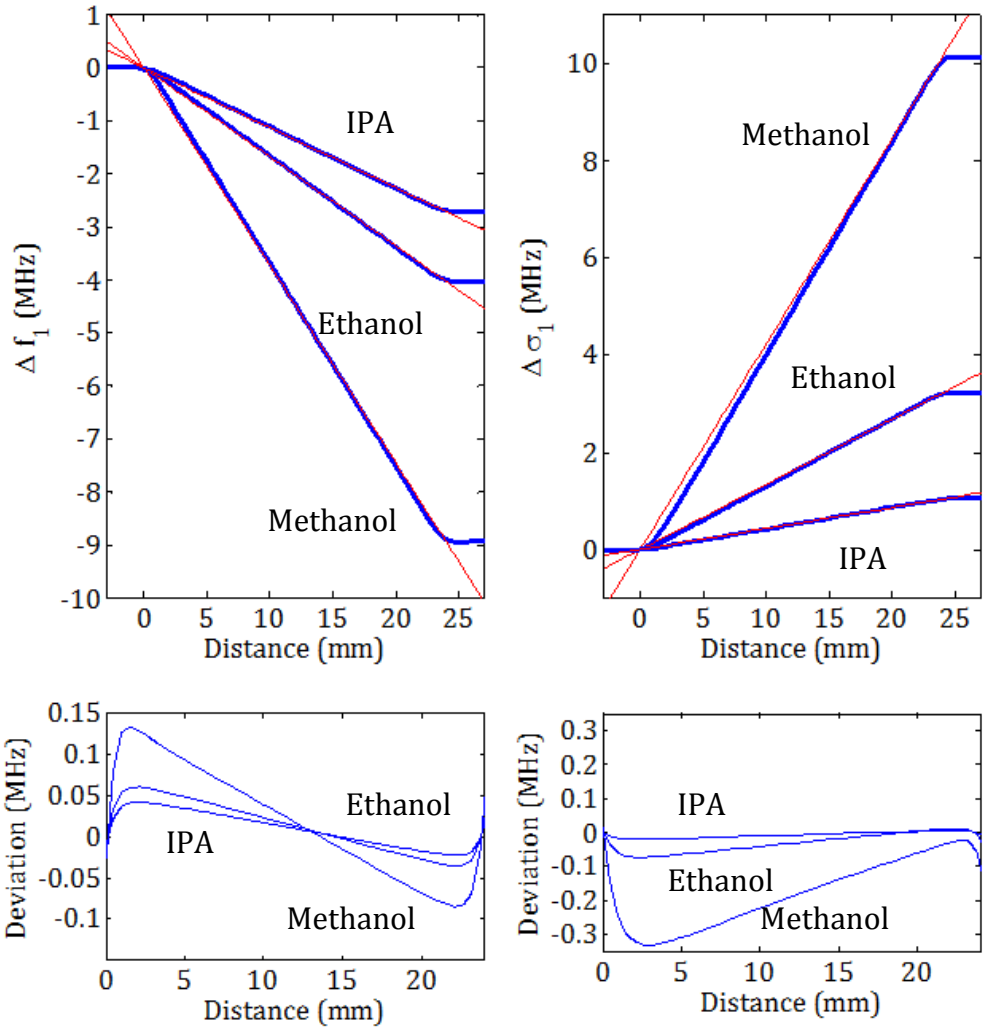



*Figure 6.36: Results from parametric FEM simulation of the capillary filling for the three different liquids, with the predicted change in resonant frequency (left) and bandwidth (right) on the top row, and the deviation of these from the linear theory (Equation 6.37) on the bottom row*

Various options were investigated for filling the capillary in a controlled manner. It was found that a high degree of variability in the time-domain filling characteristic was caused by inconsistencies in the way in which the capillary was initially brought into contact with the liquid. To mitigate this, a simple dropping mechanism was used to allow the capillary to fall from rest vertically in a repeatable way (see Figure 6.34). Another potential cause of variability was the capillary falling to the bottom of the PTFE reservoir, obstructing the free entry of fluid. A layer of polymer foam was thus used to support the capillary at a known height just below the surface of the liquid whilst providing minimal impediment to the inflow of liquid. A constant volume of liquid was pipetted into the reservoir before each experiment so that the depth of the capillary below its surface was consistent. Capillary filling begins the instant the capillary touches the liquid surface, so the distance between the stationary capillary and the reservoir liquid surface must be kept constant for the time reference for the start of filling to be the same in each case.

The apparatus was kept on a warmplate (MiniTube HT50, USA) that allowed the temperature of the resonator and reservoir to be held static. To improve heat conduction to the resonator body, a copper thermal link made from a 1 mm thick copper strip of 25 mm width was fastened to the resonator and to the warmplate with liberal amounts of heat transfer compound. Measurements were made at a range of temperatures and with various mixtures of solvents, although only a selection of these measurements will be discussed here for reasons of brevity.

### 6.4.3 Results of the capillary filling experiments

The raw temporal measurements for three alcohols are shown below. The methanol, ethanol and isopropyl alcohol (IPA) used were all HPLC grade (Sigma Aldrich, UK). These capillary filling experiments were performed with dry fused quartz capillaries (VitroCom, USA) of ID 300 µm, OD 400 µm and length 100 mm. When filling experiments were repeated with pre-wetted capillaries the filling time was found to be different. This is to be expected. Some adsorption of the alcohol to

the capillary surface will change its wettability, thereby leading to a different contact angle, as discussed in [46].

Whilst the tolerance on the inner diameter is stated as $\pm$ 10 %, the uncertainty of the sample volume can be eliminated from the dynamic phase of the measurement by referencing to difference in perturbation between empty and full (Equation 6.36). This cancels the effect of variable capillary radius under the assumption of a uniform electric field over the sample volume. Furthermore, by the standard perturbation theory, which is comparatively accurate due to the small opening in the end plate walls, the permittivity of the liquid can be estimated.

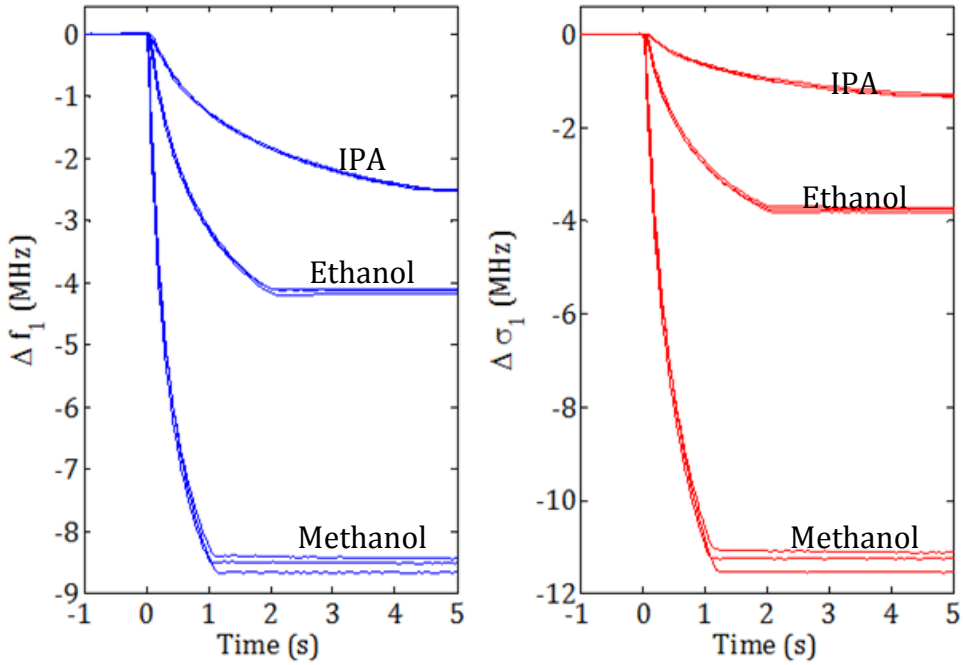



*Figure 6.37: Raw capillary filing response in resonant frequency and bandwidth for three different alcohols.*

The FEM simulations can be seen to agree quite well with the final perturbations produced by the full capillary. The results for methanol were slightly underestimated by the FEM simulation, although this is likely to be due to the difference in temperature between the actual liquid measured and the nominal 25°C value used for the simulation. To use the FEM results to convert these perturbation measurements to distance, both were first normalised to lie in the

range [0,1]. This eliminated error due to the disagreement in the absolute values of the perturbation. The simulation data were then used as a 'look up' function relating this normalised perturbation to distance. Piecewise cubic interpolation was used to match the measured perturbations with the simulated values in order to convert to distance along the resonator axis. Curve fitting to the model in Equation 6.39 was then performed on these corrected data. The results are shown in Figure 6.38 for resonant frequency (similar results were obtained for bandwidth):

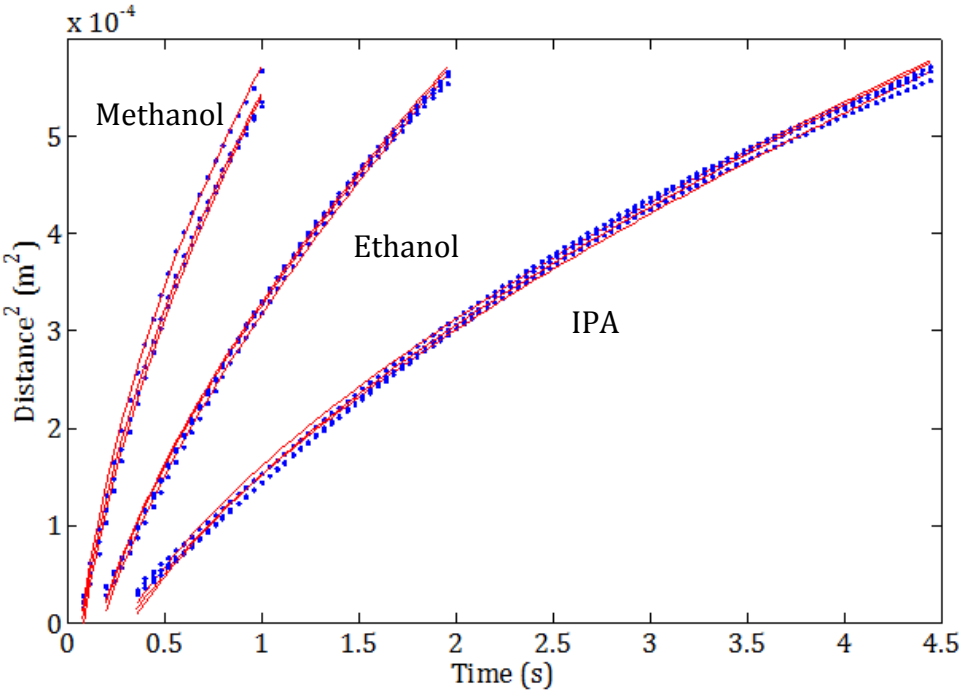



*Figure 6.38: Distance vs. time relationship of the capillary filling experiment extracted using interpolation of numerical FEM simulations. The curve fit to the Washburn equation is shown as a red curve, individual data by blue points.*

The agreement of these data with the Washburn prediction shows that dynamic capillary filling behaviour can be successfully extracted from temporal MRS measurement. However, derived estimates for the liquid diffusivity constants were not reliable, possibly due to the ambiguity of having both unknown start displacement and unknown start time for the beginning of filling making the curve-fit problem underdetermined.

## 6.5 Chapter references


[1] D. Slocombe, *Personal communication*, 2010.

[2] B. A. Dilmohamud, J. Seeneevassen, S. D. D. V. Rughooputh, and P. Ramasami, "Surface tension and related thermodynamic parameters of alcohols using the Traube stalagmometer," *European Journal of Physics*, vol. 26, no. 6, pp. 1079-1084, Nov. 2005.

[3] C. Molina, L. Victoria, and A. Arenas, "Measuring the surface tension of a liquid-gas interface by automatic stalagmometer," *Review of Scientific Instruments*, vol. 71, no. 6, pp. 2481-2486, 2000.

[4] J. Kloubek, "Measurement of the dynamic surface tension by the drop weighing method," *Colloid and Polymer Science Kolloid Zeitschrift & Zeitschrift für Polymere*, vol. 253, no. 11, pp. 929-936, Nov. 1975.

[5] A. Rothert, R. Richter, and I. Rehberg, "Formation of a drop: viscosity dependence of three flow regimes," *New Journal of Physics*, vol. 5, pp. 59-59, Jun. 2003.

[6] "Terminal Velocity." [Online]. Available: http://www.grc.nasa.gov/WWW/K-12/airplane/termv.html. [Accessed: 12-Sep-2011].

[7] "Manuals by KD Scientific." [Online]. Available: http://www.kdscientific.com/technical-resources/manuals.asp. [Accessed: 20-Sep-2011].

[8] D. M. Pozar, *Microwave engineering*. J. Wiley, 2005.

[9] A. Porch, *Personal communication*, 2010.

[10] A. P. Gregory and R. N. Clarke, "Tables of the complex permittivity of dielectric reference liquids at frequencies up to 5 GHz." National Physical Laboratory, Mar-2009.

[11] "Surface tension values of some common test liquids for surface energy analysis." [Online]. Available: http://www.surface-tension.de/. [Accessed: 13-Sep-2011].

[12] "1mm Diameter Grade 100 Hardened AISI 52100 Chrome Steel Ball Bearings Simply Bearings Ltd." [Online]. Available: http://simplybearings.co.uk. [Accessed: 11-Sep-2011].

[13] M. Perez, Y. Brechet, L. Salvo, M. Papoular, and M. Suery, "Oscillation of liquid drops under gravity: Influence of shape on the resonance frequency," *Europhysics Letters (EPL)*, vol. 47, pp. 189-195, Jul. 1999.

[14] L. Rayleigh, "On the Capillary Phenomena of Jets," *Proceedings of the Royal Society of London*, vol. 29, no. 196-199, pp. 71 -97, Jan. 1879.

[15] T. G. Wang, A. V. Anilkumar, and C. P. Lee, "Oscillations of Liquid Drops: Results from USML-1 Experiments in Space," *Journal of Fluid Mechanics*, vol. 308, pp. 1-14, 1996.

[16] A. Masood, A. Porch, and D. Barrow, *Microwave Resonators for Highly Sensitive Compositional Analysis*. Lambert Academic Publishing, 2010.

[17] Triplett K.A., Ghiaasiaan S.M., Abdel-Khalik S.I., and Sadowski D.L., "Gas-liquid two-phase flow in microchannels Part I: two-phase flow patterns," *International Journal of Multiphase Flow*, vol. 25, no. 3, pp. 377-394, 1999.

[18] K. A. Triplett, S. M. Ghiaasiaan, S. I. Abdel-Khalik, A. LeMouel, and B. N. McCord, "Gasâ€"liquid two-phase flow in microchannels: Part II: void fraction and pressure drop," *International Journal of Multiphase Flow*, vol. 25, no. 3, pp. 395-410, Apr. 1999.

[19] H. Pedersen and C. Horvath, "Axial dispersion in a segmented gas-liquid flow," *Industrial & Engineering Chemistry Fundamentals*, vol. 20, no. 3, pp. 181-186, 1981.

[20] H. Song, J. D. Tice, and R. F. Ismagilov, "A Microfluidic System for Controlling Reaction Networks in Time," *Angewandte Chemie International Edition*, vol. 42, no. 7, pp. 768-772, Feb. 2003.

[21] N. de Mas, A. Günther, T. Kraus, M. A. Schmidt, and K. F. Jensen, "Scaled-Out Multilayer Gas–Liquid Microreactor with Integrated Velocimetry Sensors," *Industrial & Engineering Chemistry Research*, vol. 44, no. 24, pp. 8997-9013, Nov. 2005.

[22] A. J. Petro and C. P. Smyth, "Microwave Absorption and Molecular Structure in Liquids. XX. Dielectric Relaxation Times and Molecular Shapes of Some Substituted Benzenes and Pyridines1-3," *Journal of the American Chemical Society*, vol. 79, no. 23, pp. 6142-6147, Dec. 1957.

[23] "Unique Properties of DuPont Teflon AF." [Online]. Available: http://www2.dupont.com/Teflon_Industrial/en_US/products/product_by_name/teflon_af/properties.html. [Accessed: 18-Sep-2011].

[24] "Toluene MSDS." [Online]. Available: http://fscimage.fishersci.com/msds/23590.htm. [Accessed: 18-Sep-2011].

[25] A. Parkash, J. K. Vaid, and A. Mansingh, "Measurement of Dielectric Parameters at Microwave Frequencies by Cavity-Perturbation Technique," *IEEE Transactions on Microwave Theory and Techniques*, vol. 27, no. 9, pp. 791- 795, Sep. 1979.

[26] "Microwave Synthesis - An Introduction," pp. 1-9.

[27] P. LidstrÃ¶m, J. Tierney, B. Wathey, and J. Westman, "Microwave assisted organic synthesisâ€"a review," *Tetrahedron*, vol. 57, no. 45, pp. 9225-9283, Nov. 2001.

[28] A. R. Katritzky, S. M. Allin, and M. Siskin, "Aquathermolysis: Reactions of Organic Compounds with Superheated Water," *Accounts of Chemical Research*, vol. 29, no. 8, pp. 399-406, Jan. 1996.

[29] "CEM Corporation." [Online]. Available: http://cem.com/. [Accessed: 15-Sep-2011].

[30] "Products :: Anton-Paar.com." [Online]. Available: http://www.anton-paar.com/Products/2_UK_en. [Accessed: 15-Sep-2011].

[31] J. J. Shah et al., "Microwave dielectric heating of fluids in an integrated microfluidic device," *Journal of Micromechanics and Microengineering*, vol. 17, no. 11, pp. 2224-2230, 2007.

[32] D. Issadore, K. J. Humphry, K. A. Brown, L. Sandberg, D. A. Weitz, and R. M. Westervelt, "Microwave dielectric heating of drops in microfluidic devices," *Lab on a Chip*, vol. 9, p. 1701, 2009.

[33] J. J. Shah, J. Geist, and M. Gaitan, "Microwave-induced adjustable nonlinear temperature gradients in microfluidic devices," *Journal of Micromechanics and Microengineering*, vol. 20, p. 105025, Oct. 2010.

[34] K. J. Shaw et al., "Rapid PCR amplification using a microfluidic device with integrated microwave heating and air impingement cooling," *Lab on a Chip*, vol. 10, p. 1725, 2010.

[35] H. Caglayan et al., "Ultrafast and sensitive bioassay using split ring resonator structures and microwave heating," *Applied Physics Letters*, vol. 97, p. 093701, 2010.

[36] E. Comer and M. G. Organ, "A microcapillary system for simultaneous, parallel microwave-assisted synthesis," *Chemistry (Weinheim an Der Bergstrasse, Germany)*, vol. 11, no. 24, pp. 7223-7227, Dec. 2005.

[37] I. R. Baxendale, J. J. Hayward, and S. V. Ley, "Microwave reactions under continuous flow conditions," *Combinatorial Chemistry & High Throughput Screening*, vol. 10, no. 10, pp. 802-836, Dec. 2007.

[38] A. J. deMello, "Control and detection of chemical reactions in microfluidic systems," *Nature*, vol. 442, pp. 394-402, Jul. 2006.

[39] C. O. Kappe and D. Dallinger, "The impact of microwave synthesis on drug discovery," *Nat Rev Drug Discov*, vol. 5, no. 1, pp. 51-63, Jan. 2006.

[40] C. S. Eskilsson and E. Björklund, "Analytical-scale microwave-assisted extraction," *Journal of Chromatography. A*, vol. 902, no. 1, pp. 227-250, Dec. 2000.

[41] Y. Zhang, R. Liu, Y. Hu, and G. Li, "Microwave heating in preparation of magnetic molecularly imprinted polymer beads for trace triazines analysis in complicated samples," *Analytical Chemistry*, vol. 81, no. 3, pp. 967-976, Feb. 2009.

[42] K. Aslan and C. D. Geddes, "Microwave-Accelerated Metal-Enhanced Fluorescence: Platform Technology for Ultrafast and Ultrabright Assays," *Analytical Chemistry*, vol. 77, no. 24, pp. 8057-8067, Dec. 2005.

[43] A. Kempitiya, D. A. Borca-Tasciuc, H. S. Mohamed, and M. M. Hella, "Localized microwave heating in microwells for parallel DNA amplification applications," *Applied Physics Letters*, vol. 94, p. 064106, 2009.

[44] J. Naylon et al., "Efficient microwave heating and dielectric characterisation of microfluidic systems," in *Proceedings of MicroTAS*, 2010.

[45] C. Gabriel, S. Gabriel, E. H. Grant, E. H. Grant, B. S. J. Halstead, and D. Michael P. Mingos, "Dielectric parameters relevant to microwave dielectric heating," *Chemical Society Reviews*, vol. 27, p. 213, 1998.

[46] "Polyflon Company - Microwave Substrates - Microwave Circuit Manufacturing, PTFE, Cuflon, NuCLAD, ULTEM, POLYGUIDE," [Online]. Available: http://www.polyflon.com/microw.htm. [Accessed: 19-Sep-2011].

[47] A. Morgan, *Personal communication*, 2011.

[48] C. Rønne, L. Thrane, P.-O. Åstrand, A. Wallqvist, K. V. Mikkelsen, and S. R. Keiding, "Investigation of the temperature dependence of dielectric relaxation in liquid water by THz reflection spectroscopy and molecular dynamics simulation," *The Journal of Chemical Physics*, vol. 107, p. 5319, 1997.

[49] U. Kaatze, "Complex permittivity of water as a function of frequency and temperature," *Journal of Chemical & Engineering Data*, vol. 34, no. 4, pp. 371-374, Oct. 1989.

[50] W. J. Ellison, "Permittivity of Pure Water, at Standard Atmospheric Pressure, over the Frequency Range 0–25 THz and the Temperature Range 0–100 °C," *Journal of Physical and Chemical Reference Data*, vol. 36, p. 1, 2007.

[51] J. Hyvert, "Dynamic measurements of the complex permittivity of dielectric liquids using HF methods," 2011.

[52] P. Colin F, "Planar chromatography at the turn of the century," *Journal of Chromatography A*, vol. 856, no. 1-2, pp. 399-427, Sep. 1999.

[53] P. Yager, G. J. Domingo, and J. Gerdes, "Point-of-Care Diagnostics for Global Health," *Annual Review of Biomedical Engineering*, vol. 10, pp. 107-144, Aug. 2008.

[54] X.-Z. Wu, T. Huang, Z. Liu, and J. Pawliszyn, "Whole-column imaging-detection techniques and their analytical applications," *TrAC Trends in Analytical Chemistry*, vol. 24, no. 5, pp. 369-382, May 2005.

[55] D. Homentcovschi, J. Geer, and T. Singler, "Uniform asymptotic solutions for small and large sessile drops," *Acta Mechanica*, vol. 128, pp. 141-171, Sep. 1998.

[56] G. Kumar and K. N. Prabhu, "Review of non-reactive and reactive wetting of liquids on surfaces," *Advances in Colloid and Interface Science*, vol. 133, no. 2, pp. 61-89, Jun. 2007.

[57] T. Young, "An Essay on the Cohesion of Fluids," *Philosophical Transactions of the Royal Society of London*, vol. 95, pp. 65 -87, Jan. 1805.

[58] P. G. de Gennes, "Wetting: statics and dynamics," *Reviews of Modern Physics*, vol. 57, no. 3, p. 827, Jul. 1985.

[59] B. V. Zhmud, F. Tiberg, and K. Hallstensson, "Dynamics of Capillary Rise," *Journal of Colloid and Interface Science*, vol. 228, no. 2, pp. 263-269, Aug. 2000.

[60] G. Martic, F. Gentner, D. Seveno, D. Coulon, J. De Coninck, and T. D. Blake, "A Molecular Dynamics Simulation of Capillary Imbibition," *Langmuir*, vol. 18, no. 21, pp. 7971-7976, Oct. 2002.

[61] R. Lucas, "Ueber das Zeitgesetz des kapillaren Aufstiegs von Flüssigkeiten," *Kolloid-Zeitschrift*, vol. 23, pp. 15-22, Jul. 1918.

[62] E. W. Washburn, "The Dynamics of Capillary Flow," *Physical Review*, vol. 17, no. 3, p. 273, Mar. 1921.

[63] G. Martic, F. Gentner, D. Seveno, D. Coulon, J. De Coninck, and T. D. Blake, "A Molecular Dynamics Simulation of Capillary Imbibition," *Langmuir*, vol. 18, no. 21, pp. 7971-7976, Oct. 2002.

# 7 Conclusions

This thesis investigated new and improved techniques for implementing microwave resonant sensor (MRS) systems. The use of microwave resonators to make high sensitivity sensors has been demonstrated in a number of challenging, dynamic applications recently, including kinetic inductance detectors, scanning microwave microscopes and miniaturised microfluidic systems. The unmet demands of high-speed and high-performance MRS measurement gave this work a clear and timely aim: to solve the issues currently limiting the speed, accuracy, and integration of MRS systems so as to enable their advantages to be brought to bear in new and unforeseen applications.

This chapter first summarises the achievements of this work, before conducting a review of the limitations and more immediate future improvements that could be made to overcome them. Finally, it concludes by suggesting some of the intriguing future developments that may arise from the contributions made herein, in the light of current trends in high-performance MRS measurement.

## 7.1 Achievements and contributions

The concept of dynamic MRS measurement has been analysed theoretically, showing that there is no fundamental reason why the rate at which temporal resonance perturbation is measured could not be comparable with the bandwidth of the resonator itself – over 100 KHz for typical sensors. Sources of nonlinearity and other inaccuracies have been studied, outlining ways to maximise accuracy without compromising sensitivity through thorough sensor modelling. This knowledge will assist in the design of future resonant sensors.

A flexible, wide-band and high-speed multitone readout system for MRS was designed and realised. Since starting this work, the concept of a multitone readout system was found to have been developed independently by Hermann *et al.,* [1]. However, the present work was unique in that it sought to maintain accuracy without compromising on speed, and was targeted at cutting-edge applications rather than conventional industrial measurement. This was ensured through applying a rigorous, all-encompassing approach using a number of novel techniques such as digital modulation, asymmetric tone spacing, and digital predistortion methods to maximise the achievable accuracy of the system. This was combined with new signal processing and sensor characterisation methods, creating a readout system with the performance necessary for the demanding MRS applications of tomorrow.

The multitone system was verified for both static and dynamic performance. Static accuracy was found to be comparable to the best readout technique available, this being a calibrated VNA with 8-term curve fitting. Long-term stability measurements identified that the noise floor of the measurement system was reached on time scales shorter than about 5 s. In this region, the dynamic range was over 180 dB (at 95 % confidence) for resonant frequency averaged over 1 s. This translates to a dynamic range of 150 dB in measurements derived from resonant frequency perturbation, outperforming many other sensor types. The effect of LO phase noise was also ruled out as a limiting factor by comparing results referenced to a high-stability rubidium oscillator. This alleviates fears that

implementations of the system with lower cost microwave oscillators would unduly sacrifice precision.

Dynamic performance was verified through a series of experiments designed to create a known perturbation under controlled time-domain variation. The first experiment measured falling liquid droplets. Standard perturbation theory proved to be in good agreement with the measured results on average. However, an unexpected phenomenon was observed; high frequency oscillations due to droplet transient dynamic response to the drop formation impulse. This is the first time such oscillations have been measured using a microwave technique. The agreement of the droplet self-resonant frequency with theory gives confidence in the correct operation of the multitone system, and is testament to the power of temporal domain resonator perturbation processing.

The second experiment measured a gas-segmented flow stream. This challenging, high slew-rate perturbation presented a rate-of-change in resonant frequency of around 2 GHz/s. It was largely successful as a proof-of-concept experiment, and we believe that it is the first time that dynamic MRS measurement at such high slew rates has been demonstrated. In addition, the system was shown to have comparable accuracy in estimating the real part of the permittivity of a low-loss dielectric liquid to that achievable with static VNA techniques, limited by the fabrication of the sensor itself. Indeed, the inversion of measured perturbation to permittivity for this sensor provided a good demonstration of the nonparametric FEM-derived sensor inverse modelling technique pioneered in this work.

The results predicted by FEM simulation of the SRR agreed more closely with the measurements of both the multitone system and static VNA. This indicates that the numerical modelling approach is superior and ultimately more flexible than the traditional analytical perturbation approach, the unwieldy expressions of which nevertheless failed to take into account fringing field effects and therefore underestimated the true resonator effective volume. With complex resonant sensor designs, the time and effort needed to derive a fully representative analytical model soon exceeds that needed to produce a complete parametric FEM

characterisation, especially given the power of modern computer workstations. This approach should be seriously considered as a replacement for perturbation theory in all but the simplest sensor geometries.

The static noise measurements showed that low frequency noise degrades the dynamic range of MRS systems below 0.2 Hz. The falling droplet experiment also demonstrated the ability of an online measurement system to subtract a periodically sampled baseline. This is of great practical significance when the environment is prone to drift (due to temperature and humidity swings, *etc*.), which would otherwise be a significant source of error in uncontrolled sensor environments, such as industrial settings or outdoors. Taken together, these findings show that both maximum precision and improved accuracy are achieved by temporal processing of a modulated measurand signal, *e.g.,* alternating between the desired sample and a reference, at frequency over 1 Hz. This finding is applicable to all microwave materials measurements. Although this concept is common, for example in astronomical telescopes to remove the background brightness [2], its application to microwave materials measurement has not been explicitly demonstrated *per se*, despite acknowledgement of the drift problem and the existence of similar 'sample chopping' apparatus to ensure *in situ* perturbation as in [3].

The multitone system was applied to the time-domain measurement of liquid complex permittivity changes induced by microwave heating. Whilst it has been known for some time that a microwave resonant sensor can also be used to heat a sample as well as measure it, this is the first time resonant sensor information has been used as a control method to create arbitrary thermal cycles in a miniaturised flow system. This is a step away from a self-monitoring microfluidic microwave reactor. This project also showed how such a system could be optimised, for both heating rate and energy efficiency, by proper design of the cavity, sample orientation, coupling control and microfluidic integration. Due to the non-depolarising geometry and high thermal isolation, heating rates were close to the maximum possible with any heating method with high loss factor liquids such as water and methanol.

The power of realtime temporal domain MRS processing was also demonstrated in a second novel application, the measurement of transient capillary filling. These experiments showed that, with the help of parametric FEM sensor modelling, it was possible to use resonator perturbation to measure both the moving meniscus position and complex permittivity of a nominally homogenous, uniform sample simultaneously. Like the falling droplet experiment, the extra temporal information made possible with realtime resonance sampling could be used to determine mechanical liquid properties as well as electrical – indicating the multi-dimensional measurement capability of high-speed, high accuracy MRS systems.

In summary, much progress has been made toward improving the speed, accuracy, and integration of MRS systems by advances in multiple directions, along theoretical, signal processing and hardware paths. The multitone readout technology can measure as fast as resonant sensors are capable of responding, and the accuracy demonstrated in certain measurements is limited only by the sensor itself, rather than the hardware, modelling or processing used. Integration remains an open challenge, and although semi-integrated solutions are presented in *e.g.,* [1] and [4], with the technology developed here, coupled with the exciting new applications and the funding opportunities they bring, it is surely only a matter of time before a truly integrated, high-performance single chip solution to MRS readout systems is possible. New and unforeseen applications have indeed been demonstrated at a proof-of-concept level, from droplet oscillations as a route to liquid rheology through to a self-monitoring microwave microfluidic reactor. For the most part, therefore, the aim of this thesis has been met successfully.

## 7.2 Limitations and improvements

The speed of the multitone system is limited to about 500 Hz for realtime measurements due to the data transfer rate bottleneck, although in single-shot mode the measurement rate is still demonstrably high (100M/256). To allow this high measurement rate to be sustained indefinitely requires the processing to be implemented on an FGPA – the logical next step in this work.

The adaptive filter and neural network algorithms are ideal for this as they allow maximum rate readout and a flexible response time. They also allow for a tuneable ratio of filtering before and after inversion to resonant frequency and bandwidth, providing optimum precision. However, they have not been tested practically under dynamic measurement conditions (as the FFT/curve-fitting algorithms have), so they are not fully verified methods and may require alteration and efficiency improvement in the light of real-world testing. One potential improvement in the neural network algorithm that might be necessary for its realisation on FPGA is the replacement of the hyperbolic tangent activation function with one more suitable for implementation with limited computational resources, such as a look-up table or piecewise linear approximation. This will require re-training of the neural network with such functions and a re-assessment of network performance.

Despite the work done on software predistortion, the system is still limited in accuracy by the intermodulation distortion, caused largely in the quadrature mixers. Distortion means that the true amplitude of the excitation tones cannot be known perfectly, making the spectral estimate, and resulting eigenvalue estimates, less accurate. Furthermore, distortion in the receiver cannot be fully corrected for by the predistortion algorithm, and the resulting spectral errors may vary as the sensor spectral response changes. This will cause a further source of perturbation-dependent error. Whilst more elaborate calibration routines, *e.g.*, at multiple attenuation levels, might improve this to some extent, a more direct solution would be to tackle the source of the distortion itself.

Some investigation into a custom passive mixer design would therefore be valuable; concentrating on maximising linearity over the frequency range of a specific sensor would undoubtedly provide better results than using a general purpose, wideband component. Furthermore, using an alternative balanced architecture to ensure excellent impedance match at the RF port would ensure minimum error due to source and load reflections. I do not believe that moving to an active mixer architecture would be a step forward. However, ways to integrate low-distortion, high dynamic range passive mixers into a system-on-a-chip or even single-chip ASIC solution would need to be investigated for the promise of true MRS system integration to be realised.

The split ring sensor measurements had some accuracy problems related to the small number of tones that occupied the resonator bandwidth at any one time. This meant that dynamic errors were introduced, especially into the bandwidth measurements, which were visibly distorted. Use of a variable gain amplifier (VGA) at the receiver would have given sufficient dynamic range for more tones to be used, largely eliminating this source of bandwidth estimate error without adversely affecting precision. The use of a VGA would also ensure optimum use of the full-scale range (FSR) of the receiver ADC at all times, making the dynamic range constant under changing bandwidth conditions and hence reducing heteroscedasticity. However, system performance would need to be re-evaluated after this system change to check for undesirable dynamic effects caused by the amplifier gain control loop. This will necessitate some further work.

The absolute accuracy of the split ring resonant sensor used for the gas segmented flow experiment is significantly worse than that of a cavity resonator, having a high sensitivity to the geometric uncertainty of small features such as the gap size and the capillary internal dimensions. It also has a depolarising geometry, restricting it to non-polar liquids. Improving the fabrication method of the SRR could increase accuracy, and the linearity of its perturbation response might be increased, *e.g.,* by using a higher permittivity capillary, or one of a different ID/OD ratio. For physical reasons, it is probably impossible to achieve high field localisation without some degree of depolarisation, therefore techniques to improve or trade-off

perturbation nonlinearity against field localisation are required; this is a good avenue for further study using parametric FEM simulation coupled with numeric optimisation techniques.

The gas-segmentation principle has promise even in otherwise static liquid dielectric measurement as a way to correct for drift automatically. However, currently it does not have the same metrological potential as the falling droplet measurement system above due to the liquid film retention problem. This caused the apparent precision of the permittivity estimate to be degraded even though the measurement system noise floor was still very low. One potential solution is to use superhydrophobic coatings on the inside of the capillary in order to suppress or eliminate liquid film retention and provide a cleaner transition from liquid to gas segment. It is also worth investigating the impact of liquid evaporation into the air segment. Although this should have minimal effect, this can quite easily be verified by performing the experiment at a range of temperatures to reach different vapour pressures. Careful temperature calibration of the resonator will be necessary, but it should be possible to show that the gas segment does not change permittivity significantly with temperature if this hypothesis holds.

The microwave heating system suffered from a number of limitations. Parasitic heating limited the accuracy of the permittivity estimate. To resolve this issue the flow system should be modified to introduce a heat exchanger, possibly located coaxially within the resonant sensor. Providing a known rate of cooling airflow would improve accuracy by preventing parasitic heating, as well as allowing for faster cooling cycles. However, heating rate and efficiency would suffer to some extent. The high level of noise and fluctuation present in the controlled heating system must also be addressed for the system to be practical. Using the latest implementation of the multitone readout system would enhance the sampling rate and provide greater noise rejection of the heating signal, as it would be phase coherent with the carrier signal used for modulation. The control algorithm could also be improved by applying digital control design techniques. Coupled with the faster readout rate, this would undoubtedly enhance the performance of the temperature control loop enough for it to become practical.

The temperature measurement also needs to be properly verified using infrared emissivity thermometry or, better still, a thermochromic tracer [5], by which spatial temperature profiles can also be recorded. In addition, the system as it stands does not represent sufficient novelty in application to interest the microfluidic community, and the sensor is still physically bulky. Combining segmented flow and heating together using the SRR to create a miniature, chip-localised microwave sensor/actuator system would certainly be a more impressive demonstration of the capabilities of microwave-microfluidics [6]

## 7.3 Scope for future development

Once realtime MHz frequency readout rates can be realised, a welcome continuation of this work would be to verify that the transient response of resonator perturbation at these very high frequencies does indeed fit with the theory of parametric state-space analysis, although conceiving a suitably fast and predictable perturbation may be difficult. The use of lasers to create rapid thermal transients or to exploit nonlinear electro-optic effects in dielectrics could be a potential candidate for such an experiment, and such a system could then provide a useful platform for the study of such interactions.

Whilst the falling droplet experiment was intended to be simply a proof-of-concept demonstration of the multitone system and realtime MRS processing, it may have some practical applications as a method of investigating liquid properties. Apart from permittivity, this experiment can measure two independent factors – drop rate and oscillation frequency – that depend on the density and surface tension of the liquid, hence allowing these useful properties to be measured. Viscosity could also be measured if the resonator could be made taller, allowing sufficient transit time for the estimation of the droplet oscillation decay constant. This is because the characteristic decay time $e^{-t/\tau}$ of the vibrational mode is related to kinematic viscosity $\eta$ in the small damping limit: [7]

$$\frac{1}{\tau} = \frac{\eta}{r^2}(n-1)(2n+1). \tag{7.1}$$

In short, if the sampling rate was increased, and the accuracy and repeatability of droplet generation optimised, this method might be competitive with much more delicate and expensive instruments such as that demonstrated by Matsumoto et al., where lasers were used to measure falling droplet shape [8]. Evidently, the metrological application of this MRS system has potential as an avenue for further study [9]. This method has the advantage of only requiring a small sample volume, and, combined with a cost-efficient, high-speed readout system optimised for small

perturbation and high accuracy, (*e.g.,* using narrower band, 16-bit converters) it could even be a candidate for a novel laboratory bench-top instrument.

With a few improvements, it is fair to say the multitone readout system coupled with the SRR is a good solution to handing the challenging measurement of gas-segmented flow streams. This sensor has a very high sensitivity and is second only to a microwave microscope's sharpened tip for electric field localisation. However, the sensor overall is much more compact than one based on Kim's static microwave microscope [10] would be, giving it unique potential. This raises the interesting possibility of new applications in online monitoring and control of segmented flow systems. Such an MRS system could assist chromatography assays that use gas segmentation post-column to reduce axial dispersion, perhaps by adding an extra analysis dimension, or simply providing a reference measurement when solvent gradients are employed. Another potential application is in gas-liquid reactors. This split ring MRS system has the potential to measure individual segments of such reactors, to determine their composition, measure reaction progress or even perform targeted microwave heating on a slug-by-slug basis.

The system capable of simultaneous complex permittivity measurement with microwave heating also has potential. The ability to measure the impact of microwave heating with the same field distribution as used to cause it can eliminate sources of uncertainty in conventional microwave heating. In particular, it gives an indication of the volume-averaged temperature through a sample. This is something that methods such as infrared emissivity thermometry cannot do. Furthermore, if reactions are progressing, knowing the exact energy absorbed via time-integrated bandwidth monitoring and measuring temporal permittivity change could together be used to monitor changes in composition or perform rough calorimetry. This information could then be fed back directly to heating control using the demonstrated PWM method. Whilst probably not a substitute for thermometry, due to the unknowns in chemical composition and parasitic heating, the extra dimensions this capability brings might still be very useful in optimising high performance flow chemistry and related applications.

Looking farther forward, if a low-cost, high-speed MRS readout system can be realised and is coupled with this technique, a wider scope of applications can be envisaged than just microreactor research. Although the technique does not scale well to higher filling factors and larger resonators (*i.e.* to satisfy domestic microwave requirements), in principle the technique could be scaled up, for example to pharmaceutical production levels, by a parallel method – duplicating a large number of fully integrated microreactors each with intelligent microwave capability. This is, after all, the preferred scale-up method for most microreactor technologies. Furthermore, the technique can be boosted with higher microwave powers to yield even more rapid heating rates. With new, low cost LDMOS transistors now capable of delivering microwave powers over 200 W at 2.5 GHz, [11] this technology and heating method is appealing for the laboratory microwave market – perhaps finally offering a fully 21st century implementation of the Bunsen burner of the 21st century.

The capillary filling experiments, whilst only simple demonstrations, also present some highly speculative applications. In the real-world scenario of measuring unknown, complex liquids, some degree of chromatographic separation might be possible under the self-filling conditions. This would be reflected in the filling dynamics, potentially creating a unique transient filling 'fingerprint'. In these situations, decoupling capillary filling dynamics and perturbation from samples also undergoing time-domain changes or having spatial inhomogeneity, is a complex and interesting challenge. One simple solution that has some rather interesting potential, is to add a secondary sensing method to detect the liquid in the capillary, for example a linear array optical detector, creating a multi-sensor system that combines the high dynamic range and robustness of microwave sensing with the additional optical spatial information.

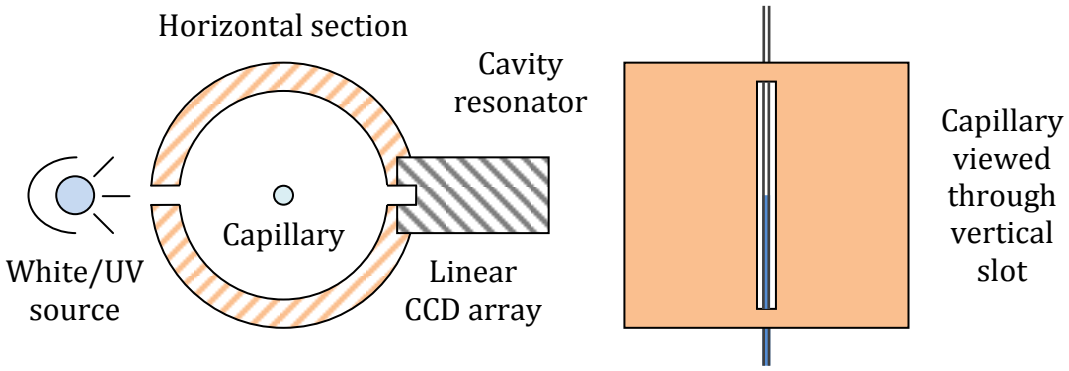



*Figure 7.1: Design for a multi-sensor dynamic capillary filling sensor integrating optical and microwave sensing.*

This multi-function sensor design uses a linear charge coupled device (CCD) operating in spectral absorption mode, with uniform illumination *e.g.,* white LED array or UV lamp for fluorescence. Linear sensors, *e.g.,* [12], have thousands of pixels, giving high spatial resolution, and can have sampling rates over 100 Hz. This multi-sensor device could replace a range of different analyses performed on complex fluids such as blood, simply by measuring complex permittivity and optical absorbance dynamically over capillary filling. Tests might consist of a selection of disposable capillaries pre-coated with a variety of compounds – such an anticoagulant agent, or an artificial receptor. The very small volume of sample needed, combined with the promise of low overall cost, with only the capillary as a disposable unit, makes this sensor system appealing for point-of-care diagnostics, hailed as enabling technology for the developing world [13].

This might seem far-fetched, and the application of MRSs to biomedical applications is indeed still in its infancy. However, recent developments in non-invasive blood glucose sensors [14], that are likely to utilise the multitone readout technology developed here, are pushing the technology of microwave sensor design and to the limits of miniaturisation, portability, and cost, this area looks set to be an exciting new area of future research.

One application area that has been alluded to several times without actually being demonstrated practically, is the microwave microscope. Using the multitone measurement system could increase the scan rate of such a device by orders of

magnitude. This could be translated to higher resolution, larger scan areas, or the ability to resolve spatiotemporally: in other words making a microwave video. In this field, the multitone system and the other approaches developed here have much to offer, and the application of the multitone system to microwave microscopy could be demonstrated in the near future simply by replacing the readout method employed on an existing microscope with the PXI-based implementation of the readout system developed in this work. As microwave microscopy is rapidly becoming a common laboratory method, there is also an opportunity here for the commercialisation of an instrument based on the PXI implementation. The concept of periodic recalibration could also be applied to the microwave microscope if the traversal velocity is quick enough to allow the measuring probe to return to a known calibration standard at the end of each scan line. This line-by-line recalibration could greatly reduce sensitivity to drift and enhance the accuracy of microwave microscope measurements.

In summary, the continued expansion of microwave resonant sensors into new disciplines such as chemistry and medicine will require great advances in their readout systems and design methods, but the rewards will be a diversity of new applications as a next-generation embedded sensor technology. It is hoped that the contributions made in this work will prove to be a valuable step towards this goal. Furthermore, high-performance realtime MRS measurement is an enabling technology in more traditional fields such as material metrology and microwave microscopy, offering new capabilities and higher performance, and aiding advancement across the scientific disciplines.

## 7.4 Chapter references


[1] T. Hermann, G. R. Olbrich, and P. Russer, "A Novel System for Real-time Measurement of the Electrical Properties of a Cavity Resonator," in *Microwave Conference, 2008. EuMC 2008. 38th European*, 2008, pp. 67-70.

[2] I. S. McLean, *Electronic imaging in astronomy: detectors and instrumentation*. Springer, 2008.

[3] S. Donovan, O. Klein, M. Dressel, K. Holczer, and G. Grüner, "Microwave cavity perturbation technique: Part II: Experimental scheme," *International Journal of Infrared and Millimeter Waves*, vol. 14, pp. 2459-2487, Dec. 1993.

[4] T. Sokoll and A. F. Jacob, "Self-calibration circuits and routines for low-cost measuring systems," *Microwave and Optical Technology Letters*, vol. 50, no. 2, pp. 287-293, Feb. 2008.

[5] D. Ross, M. Gaitan, and L. E. Locascio, "Temperature Measurement in Microfluidic Systems Using a Temperature-Dependent Fluorescent Dye," *Anal. Chem.*, vol. 73, no. 17, pp. 4117-4123, Sep. 2011.

[6] J. Naylon, "Realtime microwave resonant sensor readout system for flow chemistry applications [to be published]," *Sensors and Actuators A: Physical*, Expected 2012.

[7] A. Rothert, R. Richter, and I. Rehberg, "Formation of a drop: viscosity dependence of three flow regimes," *New Journal of Physics*, vol. 5, pp. 59-59, Jun. 2003.

[8] T. Matsumoto, H. Fujii, T. Ueda, M. Kamai, and K. Nogi, "Oscillating drop method using a falling droplet," *Review of Scientific Instruments*, vol. 75, p. 1219, 2004.

[9] J. Naylon, "Measurement of liquid droplet oscillations with a realtime microwave sensing method [to be published]," *Journal of Applied Physics*, Expected 2012.

[10] J. Kim, A. Babajanyan, A. Hovsepyan, K. Lee, and B. Friedman, "Microwave dielectric resonator biosensor for aqueous glucose solution," *Review of Scientific Instruments*, vol. 79, p. 086107, 2008.

[11] "2400-2500 MHz ISM Band from NXP Semiconductors." [Online]. Available: http://www.nxp.com/#/ps/ps=[i=71640]|pp=[t=pfp,i=71640]. [Accessed: 30-Sep-2011].

[12] "ALPHALAS - Digital Linear CCD Arrays with USB Interface and LabVIEW Drivers: CCD Series." [Online]. Available: http://www.alphalas.com/products/. [Accessed: 20-Sep-2011].

[13] P. Yager, G. J. Domingo, and J. Gerdes, "Point-of-Care Diagnostics for Global Health," *Annual Review of Biomedical Engineering*, vol. 10, pp. 107-144, Aug. 2008.

[14] A. Porch and J. Beutler, "Non-invasive blood glucose meters," Wellcome Trust transfer award, 2011.