Imaging Medical Endoscopes In Three Dimensions Using Magnetic Fields

by

John Stuart Bladen BEng

A thesis presented in candidature for the degree of Doctor of Philosophy at the University of Sheffield

Department of Electronic and Electrical Engineering

September 1995

Imaging Medical Endoscopes In Three Dimensions Using Magnetic Fields

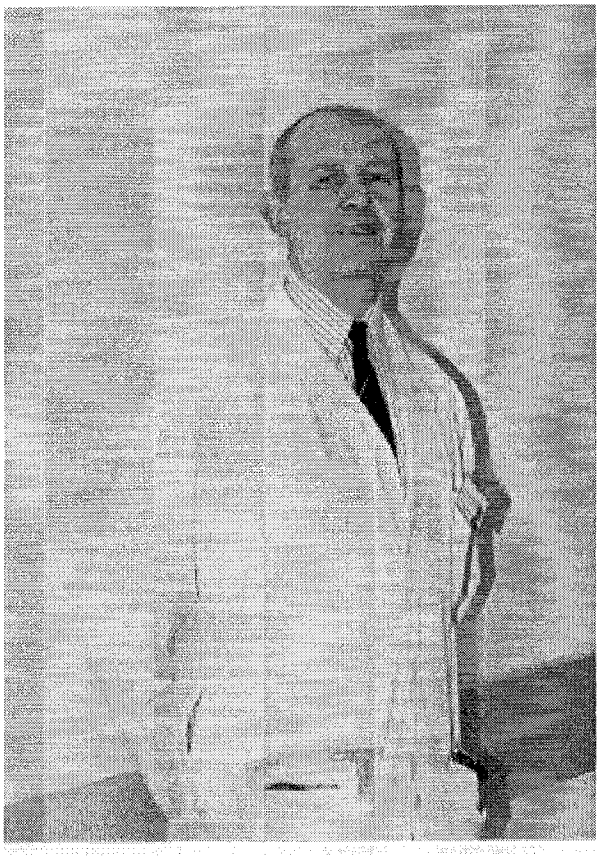
J.S. Bladen BEng

Summary

This thesis describes a novel imaging system based on very low strength magnetic fields capable of displaying the entire path of a medical colonoscope in three dimensions. The thesis begins with an overview of existing medical imaging modalities and their suitability for locating the path of a colonoscope. Having rejected these on the basis of resolution, cost or safety, the use of magnetic fields as the basis for the imaging system is established as the most promising approach compatible with clinical practice.

The proposed method involves determining the position of discrete points along the colonoscope and then reconstructing an image of the configuration of the colonoscope by fitting a 3D curve between the points. An array of miniature magnetic field sensors is inserted into the existing biopsy channel of the colonoscope and the position and orientation of each sensor is determined using a novel position location algorithm. The predicted accuracy of the system is simulated in software and the development of a practical system which implements the position location algorithm is discussed. This system comprises high performance data acquisition hardware to produce real time images of the colonoscope as it is inserted into the patient. The design of each performance critical component of the system is examined and optimised to maximise the positional accuracy of the system.

Extensive clinical trials have been performed using the colonoscope imaging system, and these are described. The trials show the effectiveness of the system as a real aid to performing colonoscopy, even for an expert endoscopist. The thesis concludes by discussing the goals met by the system and future studies that might be carried using the system, such as reducing the training time for colonoscopy. Also, other medical and non-medical applications are discussed where the position and orientation of miniature sensors is required.



Dr G D Bell, the initiator of the project

Publications Arising From The Thesis

1) J.S. Bladen, A.P. Anderson, G.D. Bell, B. Rameh and D.J.T Heatley : "Non-radiological technique for three-dimensional imaging of endoscopes", The LANCET, Vol. 341, pp 719-722, 20th March 1993.

2) J.S. Bladen, A.P. Anderson, G.D. Bell and D.J. Heatley : "A non-radiological technique for the real time imaging of endoscopes in 3 dimensions", IEEE Medical Imaging Conference, Vol 3, pp 1891-1894, Nov 1993.

3) B.P. Saunders, G.D. Bell, C.B. Williams, J.S. Bladen, A.P. Anderson : "First clinical results with a real time, electronic imager as an aid to colonoscopy", GUT, Volume 36, pp 913-917, 1995.

4) Patent: J.S. Bladen and A.P. Anderson : "Position Location System", International Publication Number: WO 94/04938, International Publication Date: 3/3/94.

5) Newspaper articles:

Independent "Virtual reality opens up hope of remote surgery", 27/1/93.
Daily Telegraph "Virtual reality creates a 3-D theatre for far-off surgeons", 27/1/93.
Yorkshire Post "Scientists pioneer surgery of future", 27/1/93.
Times "When operations on TV become the real thing", 19/3/93.

6) "Beyond 2000", Australian TV science documentary, 1994.

Glossary of Terms

Abdominal pressure	Hand pressure on adbomen of the patient undergoing colonoscopy to help insertion of the colonoscope.
Anastomosis	Join between two parts of colon which is made when a portion of the colon is surgically removed.
Anterior / Posterior	Front to back view of patient.
Caecum	Small sack at the start of the colon. Reaching the caecum indicates that the colonoscope is fully inserted.
Colon	Portion of gut between caecum and rectum.
Colonoscope	Endoscope which is optimised for examining the colon.
CPU	Central processing unit.
Eddy currents	Currents induced in conductive objects by time variant magnetic fields. These currents in turn produce magnetic fields which distort the original field.
Endoscope	Flexible tube with an optical device at the tip which is inserted into the patient and controlled from the operator end.
ENDOVIEW	Specially written program for recalling the images recorded during a patient examination.

Generator coil assembly	Assembly of three coils wound in orthogonal directions.
Ideal magnetic dipole	Theoretical current carrying coil that produces a field that obeys the magnetic dipole field equation.
Imager	Name given by clinicians to the endoscope location system described in this thesis.
Intubation	Process of inserting an endoscope into the patient.
Left lateral position	Patient lying on their left side, the standard position for colonoscopy procedure.
MAGCALC	Specially written program for simulating the imager system.
mmf	Magnetic Motive Force often called *H*.
Practical magnetic dipole	Current carrying coil that produces a field that approximates that given by the magnetic dipole field.
Phase compensation	Method of reducing positional errors due to phase shifts caused by eddy currents.
Sensor catheter	Array of sensor coils mounted inside a tube, for insertion into the biopsy channel of a colonoscope
SENSOR3D	Specially written program for calculating the effect of different designs of sensors.

System frequency	Operating frequency of system.

Contents

Chapter 1

Introduction

1.1 The Colon

"Colorectal cancer is the second most common cause of death due to malignancy in England and Wales. 25000 new diagnoses are made each year, of which 17000 will die. Colorectal cancer accounts for 10% of all cancer deaths and is responsible for 3% of all deaths in the UK. About one in twenty of the British population will get colorectal cancer and only about one-third will survive five years or more" [1.1].

A knowledge of the anatomy of the colon is important to many aspects of this work. The colon, shown in figure 1.1, is a tube of about one metre in length which forms part of the alimentary canal. It starts at the caecum, which is a pouch at the end of the small intestine, and ends at the rectum. The caecum also has the appendix attached to it. The first section of the colon is called the ascending colon, or 'right' colon as viewed from the patients point of view. The ascending colon is usually bound fairly tightly to the back wall of the abdominal cavity and runs from the caecum up the right side of the abdomen to a point under the liver called the hepatic flexure. After this bend, the transverse colon crosses the abdomen to the left side, where it takes a sharper bend called the splenic flexure. Unlike the ascending colon, the transverse colon is more mobile. The descending colon, also called the 'left' colon, is more rigidly fixed, and runs from the splenic flexure to the sigmoid colon, named after the Greek letter 'Sigma'. The sigmoid colon is again mobile and can be the longest section of the colon. After the sigmoid, the colon ends at the rectum.

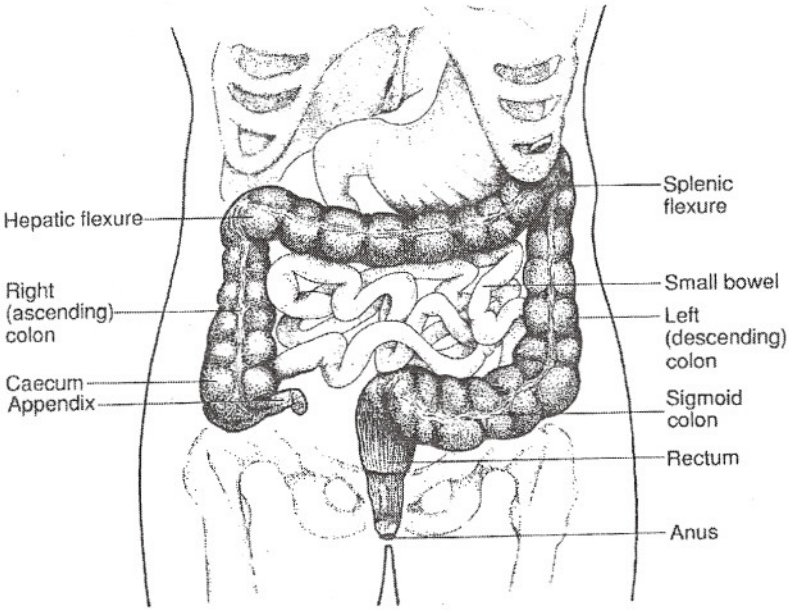


Fig 1.1: The colon - a tube of 1 metre in length which passes from the caecum to the rectum [1.2]

Colon cancers are thought to start as polyps and later develop into cancers [1.3]. Polyps are projections above the mucosal surface [1.4] and can be seen in figure 1.2. They are common, especially in older people. Around 40-60% of people are thought to have one or more polyps by the end of their life [1.5, 1.6, 1.7], though not all polyps develop into cancers. Polyps which look likely to develop further can be removed during colonoscopy.

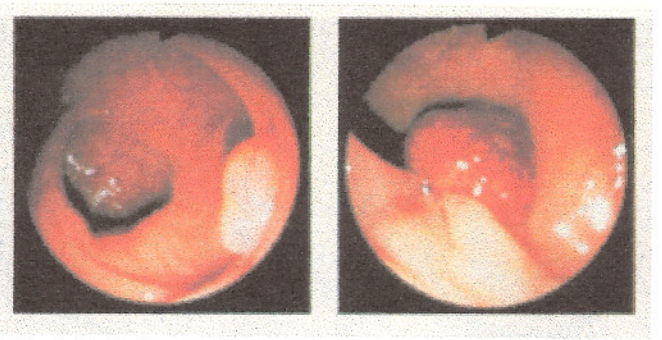


Fig 1.2: Polyps viewed from inside the colon [1.8]

1.2 Colonoscopy

As with all cancers, early detection is vital in the treatment of colorectal cancer. One way of detecting, and in some cases removing, colorectal cancers at an early stage is by using the technique of colonoscopy. Colonoscopy is the examination and treatment of the rectum, sigmoid and colon using a colonoscope and has become an important diagnosis and treatment technique in recent years in the fight against colon cancer.

Colonoscopy involves inserting a flexible tube along the colon to gain access to different parts of the colon. Typically, the colonoscope is about 1.5 cm in diameter and 1.6 m in length and features either a fibre optic system or a video camera at the tip which provides the endoscopist with a high quality image of the colon wall. By inserting the instrument along the colon, the entire colon from caecum to rectum can be examined for polyps, cancers and other abnormalities.

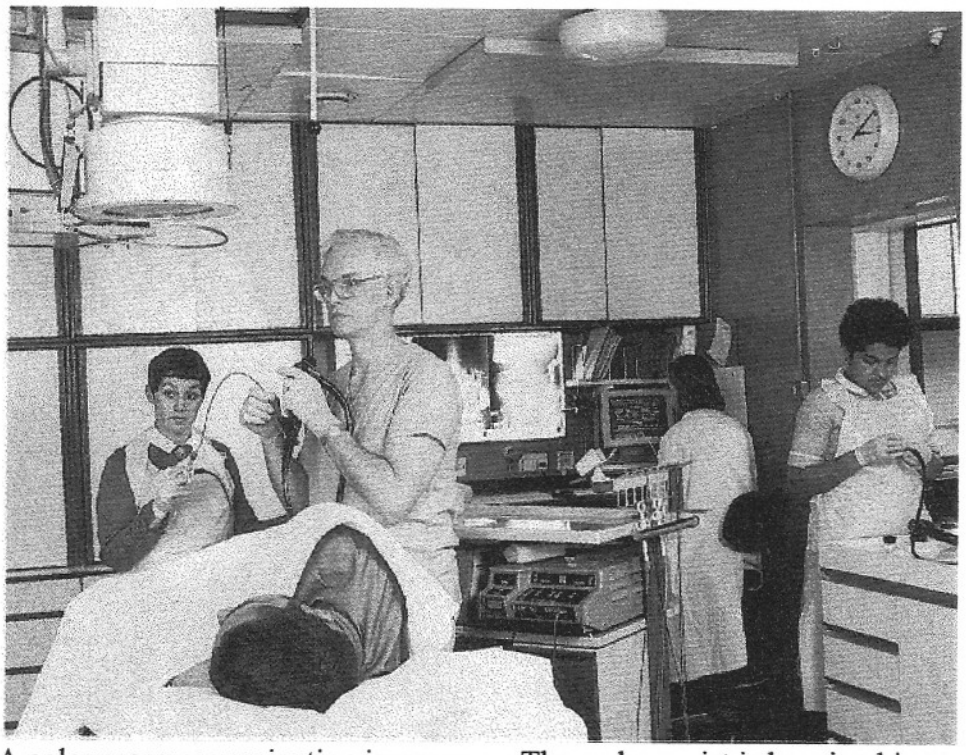


Fig 1.3: A colonoscopy examination in progress. The endoscopist is keeping his eyes on the endoscopic view which is displayed on a screen in front of him [1.2]

The colonoscope also contains a tube called the *biopsy channel*, which passes the length of the instrument, through which tools may be passed to take tissue samples or to carry out minor surgical procedures. The internal construction of the instrument is shown in figure 1.4.

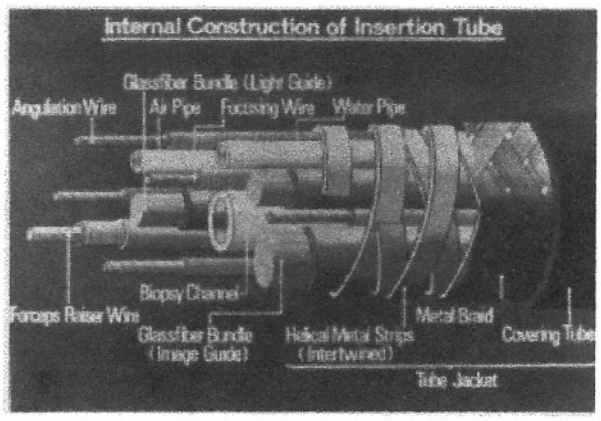


Fig 1.4: The internal construction of a colonoscope [1.9]

Polyp removal, known as polypectomy, can be carried out during colonoscopy by feeding a wire snare along the biopsy channel and looping it around the polyp. Electro-coagulation, or diathermy, is used to prevent bleeding if the polyp stalk is large. This involves passing an electrical signal of 1 MHz of up to 25 Watts between the polyp and a metal plate placed on the leg of the patient. If the polyp cannot be easily removed, then tissue samples can be taken and later examined by the histolopathologist to check for malignanancy. If a cancer is found, then its position can be confirmed prior to invasive surgery being performed to remove the cancer. The polypectomy sequence is shown in figure 1.5.

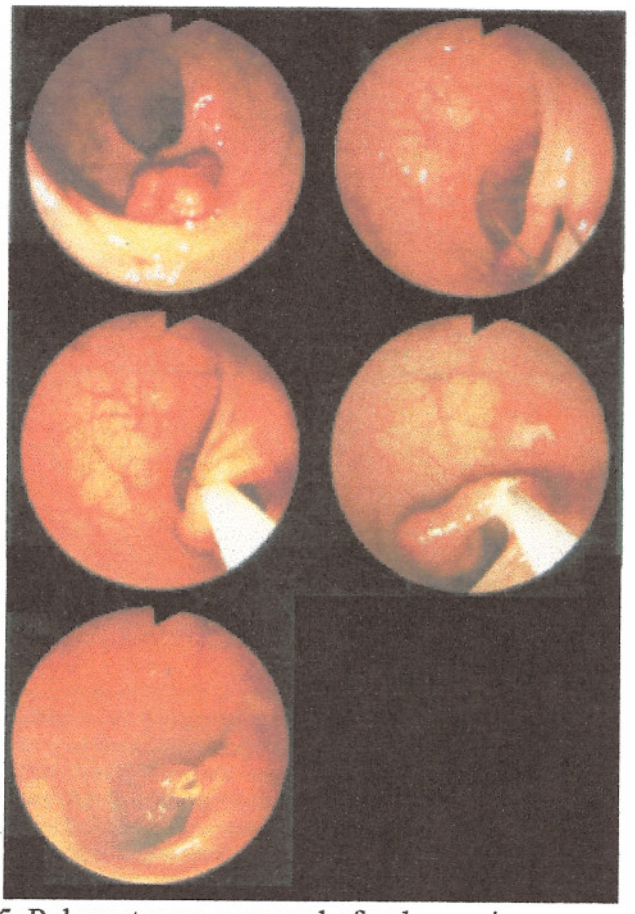


Fig 1.5: Polypectomy - removal of polyps using a snare [1.8]

Colonoscopy may also be used for many other reasons. Examination of the bowel can reveal or confirm a number of illnesses, including ulcerative colitis and Crohns disease. Also, patients with anaemia are often recommended to undergo colonoscopy to check for bleeding in the colon.

1.3 Colorectal Cancer Reduction Through Colonoscopy Screening

Widespread screening of the colon would enable polyps to be removed before they develop into cancers. Since there is little risk in removing polyps, they could be removed even if thought to be benign. Colon screening would also detect cancers in their early stages.

Colonoscopy in its current form is not suitable for such widespread screening. It is an expensive procedure and also the risks of injury become significant when performed on large numbers of people [1.2]. The high cost is primarily due to the length of the procedure. A colonoscopy examination can take a long time, up to 90 minutes, and so is tedious for both the patient and endoscopist, and makes it more than twice the cost of a gastroscopy (examination of stomach and duodenum). There is also a small risk of about 1 in 1700 patients [1.10] that perforation of the gut will occur during colonoscopy, due to the mechanical stresses of the colonoscope against the colon wall. Perforation is serious since the perforation can often become septic. For a patient who is seriously ill anyway, this risk is considered to be acceptable. However for a screening programme, where most of the patients will not be ill, this risk is considered too high.

A further problem with colonoscopy is that it is a difficult skill to acquire, often requiring a trainee endoscopist to carry out 200 supervised examinations before they can do the examination on their own [1.11]. This has resulted in there being an insufficient number of fully trained endoscopists to carry out a screening programme.

During the examination, sedatives are given to reduce the pain to the patient. The sedatives (Midazolam and Pethidine) leave the patient conscious so that they can report any pain during the procedure, but also cause the patients not to remember anything later. The sedatives mean that the patient must not drive or work after the examination. Thus the patient must have a full day off work and this increases the overall cost of carrying out the screening programme.

1.4 Solving The Problems Of Colonoscopy

The principal difficulty with carrying out the colonoscopy procedure is that of inserting the colonoscope into the caecum with reasonable comfort and safety [1.12, 1.13]. This process, known as *intubation*, frequently takes the majority of the total examination time. Since many

parts of the gut are not rigidly fixed, the colonoscope does not travel smoothly along the intestine, but rather forms loops, leaving little or no resolved force component in the forward direction (see figure 1.6). Once loops have formed, further insertion of the instrument into the patient merely causes the loop to enlarge, causes pain to the patient and achieves no advancement of the instrument tip. To remove the loop it is frequently necessary to pull the colonoscope back, coupled with simultaneous clockwise or anti-clockwise twist on the shaft of the instrument.

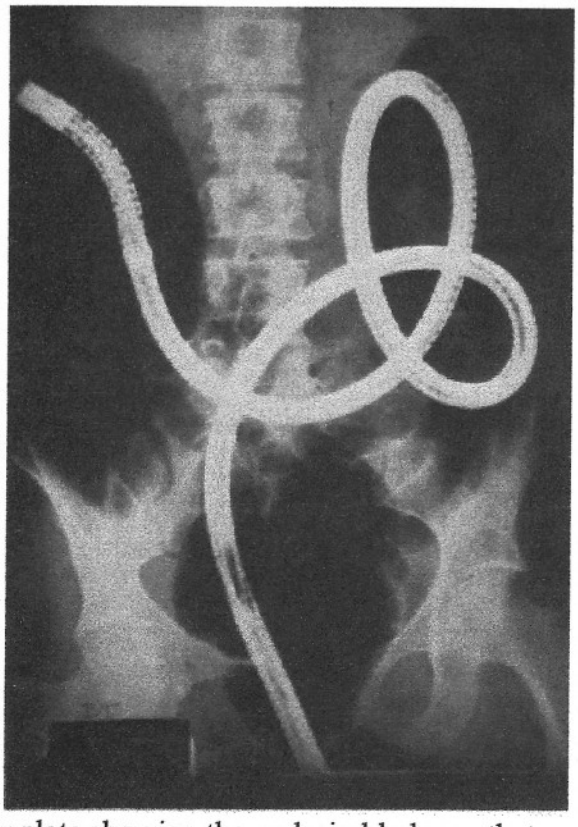


Fig 1.6: X-ray plate showing the undesirable loops that can form in the colonoscope during colonoscopy [1.14]

The endoscopist currently has no way of visualising the loops that have formed along the instrument, since the only view available is that from the tip of the instrument. In the past, fluoroscopy was used to image the instrument but was never found to be satisfactory. It is seldom used now, but it served to show that knowing the nature of the loops is a considerable help in determining the twist and pull actions required to remove them [1.15].

What is needed is a method of reducing the difficulties of intubation. If the endoscopist had a method of imaging the entire configuration of the colonoscope while it is inside the patient, then the loops could be seen forming and immediate action could be taken to remove them. Such a system would have many benefits. The procedure would become significantly shorter, since the intubation time is usually a significant fraction of the procedure time. The risk of perforation would be reduced, since the endoscopist could manage the instrument better. Also, the levels of sedatives could be reduced, since better control of the instrument will result in reduced pain to the patient. The system would also ease the learning curve for trainee endoscopists and therefore enable more doctors to become proficient at colonoscopy.

A further reason for being able to image the path of the colonoscope is to determine how much of the colon has been examined. Some countries, such as the United States, legally require the endoscopist to sign whether they have reached the caecum at the end of the colon. This is because cancers are increasingly found in the ascending colon or in the caecum [1.16]. Although there are anatomical markers for finding out whether the colonoscope tip is in the caecum, mistakes are made. The US President, Ronald Reagan, was given the 'all clear' by a endoscopist who thought he had examined the entire colon. In fact only half the colon had been examined and a cancer in the caecum had been missed. Without an imaging system, endoscopists are frequently optimistic, thinking that they have progressed further than they actually have.

1.5 Alternatives To Colonoscopy

In view of some of the difficulties of colonoscopy, other methods of inspection are sometimes used. One of these is to give a barium enema. This involves injecting a mixture of barium and air into the colon and viewing the patient using fluoroscopy (a real time imaging system which is based on x-rays). Figure 1.7 shows a typical view, in this case with a cancer in the sigmoid colon.

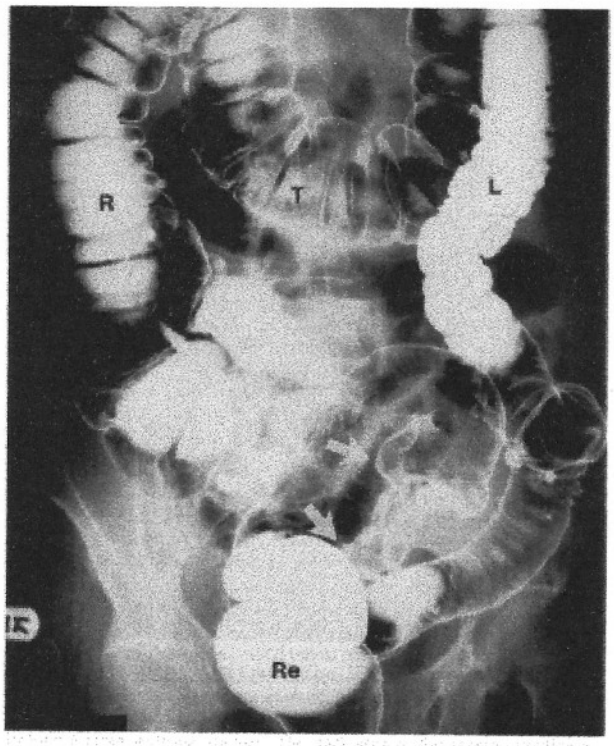


Fig 1.7: Barium enema. The arrows indicate a cancer in the sigmoid colon [1.2]

Although a barium enema is easier to carry out than colonoscopy and carries a lower risk of perforation (1 in 25000) [1.10], it is less likely to pick up the presence of polyps and lesions. Also, tissue samples cannot be taken to establish whether a polyp is benign or malignant and, of course, polyps and cancers cannot be removed..

1.6 Aim Of This Research

The aim of this research was to enhance colonoscopy by developing a system capable of imaging the entire colonoscope, making it easier to manage the instrument whilst it is inside a patient. The system would be required throughout the procedure, and therefore must be able to be run continuously without any risk to the patient or staff. The system should be real time and should be accurate enough to enable the *configuration* of the endoscope to be clearly and unambiguously seen. Absolute positional accuracy is not necessarily required. Since the loops

that form are generally in three dimensions, three dimensional imaging is preferable to two dimensional imaging. The main objective of the system is to enable better handling of the colonoscope, so it is not necessary to image the patient as well as the instrument. Of course, if it were possible to provide some registration of the endoscope with respect to the body organs this would be desirable.

1.7 Specification For An Imaging System

In conjunction with Dr Duncan Bell, the following specification for the colonoscope imaging system was devised:

- The system must be able to determine the entire path of the colonoscope, not just the tip, since loops can only be removed if the whole path is known.
- The system must image the colonoscope in 3D since the loops form in 3D.
- The system should have a resolution and accuracy of around 2 cm.
- The system should work with any colonoscope. The cost of colonoscopes is high (about £14000) and hospitals will not wish to replace all their existing instruments.
- The image must be displayed in such a way that the endoscopist can see the path of the entire colonoscope at a glance.
- The information should be displayed in real time, giving immediate feedback in changes of position of the colonoscope.
- For reasons of safety and convenience, ionising radiation (which includes x-rays) should not be used. Exposure to high power RF and other hazardous forms of radiation should also be avoided.
- No part of the imaging apparatus should hinder the endoscopist in carrying out the procedure.
- Any part of the imaging system which will come in contact with the body must be tolerant to the sterilisation procedures used in hospitals.

1.8 Survey Of Existing Imaging Modalities

Many different imaging systems are already in use in hospitals. As a starting point, these were considered to see whether they could be adapted for imaging the colonoscope. Figure 1.8 shows a table of the methods that were considered.

Imaging method	Mechanism of measurement	Physical characteristic measured	General notes
X-ray imaging	Transmission of x-rays	X-ray attenuation	2D projection of 3D distribution. Sensitivity: 2% contrast. Resolution: < 1 mm
Fluoroscopy	Transmission of x-rays	X-ray attenuation	2D imaging. Real time version of conventional x-ray imaging
X-ray computed tomography (CT)	Transmission of x-rays	X-ray attenuation	3D imaging. Resolution: 1-2 mm
Ultrasound	Reflection of ultrasonic waves	Attenuation, scattering and velocity	Spatial resolution (at 5 MHz): 5-10 mm at 50 mm depth in soft tissue. 3 mm in water
Magnetic resonance imaging (MRI)	Magnetic and electric fields	Spin relaxation times, proton density	3D imaging. Spatial resolution: < 1 mm
Planar radio isotope imaging	Nuclear particle emissions	Density of administered isotope	2D projection of 3D object distribution
Emission computed tomography (ECT) (including positron emission tomography (PET))	Nuclear particle emissions	Density of administered isotope	3D imaging. Spatial resolution: 20-40 mm
Microwave imaging	Transmission of microwaves	Microwave attenuation	2D projection of 3D distribution. Spatial resolution: 25 mm
Electrical impedance tomography (EIT)	Voltage potentials on body surface	Electrical conductivity	2D tomographic slice. Spatial resolution: 10% of diameter (~ 25 mm)
Optical imaging	Transmitted light	Light attenuation	New technique using measurement of ballistic light through tissue

Fig 1.8: Survey of existing medical imaging techniques [1.17]

X-ray imaging: X-ray imaging involves projecting a beam of x-rays through the body and detecting the transmitted x-rays. Either photographic film may be used to detect the x-rays,

producing a static image, or an x-ray camera may be used to provide real time imaging (fluoroscopy). In either case, the result is a 2D projection of the 3D distribution of the x-ray attenuating properties of the tissue. To obtain a 3D image, x-ray computed tomography (CT) is used. In CT the x-ray source and detector are mounted on a scanning system so that the whole assembly can be rotated around the patient. The image obtained from CT appears as a slice, a few millimetres thick, removed from the body. The darkness of a particular part of the image represents the electron density of the tissue, which is related principally to its physical density. The spatial resolution is about 1 mm and the density discrimination is around 1% [1.17]. CT imaging is sometimes used for imaging the colon, see figure 1.9, but is not suitable for the present objective since it prevents access to the patient during the scanning procedure. In any case, x-rays have been excluded by the specification on safety grounds.

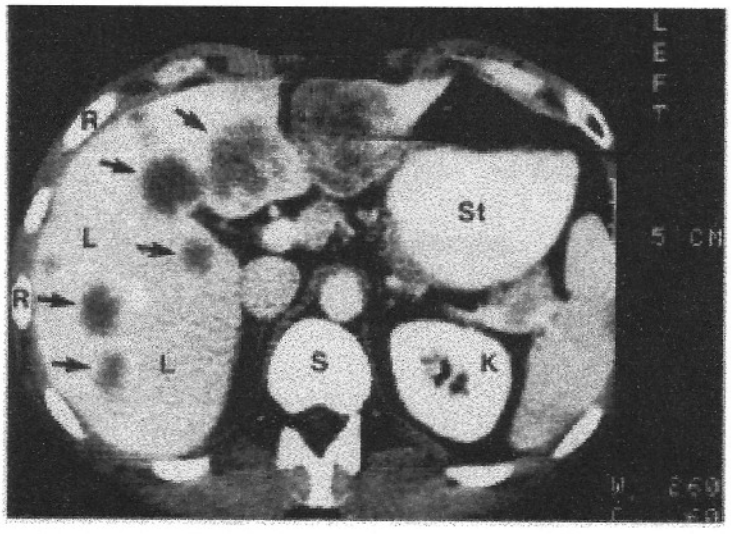


Fig 1.9: CT scan showing a tomographic slice through the lower abdomen.
The arrows shows possible cancers [1.2]

Fluoroscopy has been used in the past to image colonoscopes in the body [1.14, 1.15], but due to a number of significant problems, is rarely used these days. The area imaged by fluoroscopy is typically only 9 inches square, which is only a fraction of the size of the abdomen. To get a complete image, the endoscopist has to stop the procedure and scan the equipment over the patient manually. The image is inherently only 2D, making it difficult to visualise the complex loops which may form in the colonoscope. Fluoroscopy is also

cumbersome to use, since to minimise the x-ray exposure to both the staff and the patient, the staff have to wear heavy lead coats throughout the examination and the system cannot be used continuously. In summary, the prohibitive cost of fluoroscopic equipment and the considerable inconvenience in its use has outweighed its potential benefits of reducing intubation time and patient discomfort.

Ultrasound: Ultrasound imaging is a time of flight technique that produces a density map of the body. It uses longitudinal pressure waves in the frequency range 3 kHz to above 5 MHz, which are generated and detected at the surface of the patient. The velocity and attenuation of the ultrasound as it travels through the body depends on the tissue it is travelling through. Also, at interfaces between tissues of different densities, a proportion of the incident ultrasound is reflected. Ultrasound imaging systems use this information to produce a map of tissue boundaries and tissue densities.

There are several problems with using ultrasound for detecting the path of a colonoscope. Firstly, the colon contains gases. These have a low density compared with the surrounding tissue and so nearly all the incident ultrasound is reflected [1.18]. Thus no image of the colonoscope is possible. Secondly, ultrasound systems currently in use are naturally tomographic, meaning that they produce a 'slice' through the patient, and are therefore not suitable for obtaining a plan view of a 3D object. Thirdly, the ultrasound transmitter and receiver must be in contact with the surface of the body. Because of the way colonoscopy is currently performed, contacting transducers are a considerable inconvenience. Finally, the variation in the density of the surrounding tissue will create problems, since it will not be possible to image the colonoscope without imaging the surrounding tissue as well.

Magnetic Resonance Imaging (MRI): Magnetic resonance imaging exposes the patient within a tubular structure to a large magnetic field (usually around 1.5 Tesla) which modifies the spin of the protons in the water molecules inside the body. Magnetic resonance imaging produces an image which relates to the proton density, the freedom of hydrogen containing

molecules to rotate, and to the proportion of water contained in different body-fluid compartments.

The general principle is therefore not applicable to the location of a metal colonoscope. In any case, the large fields used by MRI could not be used in the presence of the colonoscope, which contains ferromagnetic materials.

Radio Isotope Imaging / Emission Computed Tomography: Possibilities for endoscope imaging exist here, such as mounting radio active sources along the length of the colonoscope and using radiation cameras placed around the patient to detect the position of the sources. However, this is not an attractive option in terms of safety and cost, and has not been considered in the present context.

Microwave Imaging: Two options exist for using microwave imaging to locate the colonoscope. The patient may be illuminated with microwaves which are then detected using a microwave camera. This configuration can produce a resolution of 15 to 25 mm [1.19, 1.20] using a water bolus to make microwave contact with the patient. It may be possible to image the colonoscope using this method. An alternative method is to arrange for the colonoscope to act as a microwave transmit antenna inside the body and utilise a 'microwave camera' or other external array of sensors.

A significant problem with both of these methods is that the patient must be in microwave contact with the microwave camera. This usually involves using a large water bolus and this would be a considerable inconvenience for colonoscopy. The image is only 2D and would require the imaging equipment to be rotated to obtain alternative views. A further problem with microwave imaging is that microwaves are dependent on the permittivity of the material that they are travelling through. For example, waves travel faster through the gases in the intestine than through the water in the surrounding tissue. Generally, scattering and diffraction will lead to distortion of the colonoscope image which will be difficult to remove.

Electrical Impedance Tomography (EIT): Electrical impedance imaging is a technique which determines the conductivity distribution in a slice of the body. A set of electrodes are attached to the surface of the patient, through which currents can be made to flow though the body, and surface potential differences can be measured. The information gathered can be used to determine the conductivity distribution in the plane of the electrodes. The resolution obtainable is a function of the number of electrodes and the signal to noise ratio of the system. For a system using 16 electrodes, a resolution of about 10% of the body diameter can be achieved [1.21]. This resolution is inadequate for the imaging the colonoscope. Also, the image produced is a 2D tomographic slice, which is not suitable for imaging the 3D configuration of the colonoscope.

Optical Imaging: The body is not completely opaque to light and it may be possible to emit high intensity light from the colonoscope and detect it with sensors outside the patient. The attenuation of the body to light is high and obstructions such as bone would further obscure the light. However work in the field of using optical imaging as an alternative to x-ray imaging is being carried out by several research teams [1.22] using high powered lasers. By measuring the ballistic light rather than the scattered and reflected light, an undistorted image can be obtained. The technology is still at an early stage of development.

1.9 Alternative Approaches To Imaging

It is clear that none of the standard medical imaging techniques systems described above are suitable for the present purposes. Therefore, it was necessary to look at alternative forms of imaging, specifically designed to solve the problem of imaging a colonoscope within the body during clinical examination. Some of the requirements of imaging a colonoscope are quite different from those of conventional medical imaging. One difference is having access to the object that is being imaged.

One approach which was considered, which takes advantage of this, is to insert some form of strain gauge into the biopsy channel of the colonoscope which could detect the direction and degree of bending at all points along the instrument. The strain gauge considered was a special type of fibre optic cable, where the amount of reflected light is dependent on the degree of bending, and the time delay of the reflected light is dependent on the position of the bend. To produce direction information, three cables are used, forming a triangular cross section. However, an initial study of this system did not look hopeful, due to the fact that the system relies on 'dead reckoning'. That is, all the later measurements depend on all the previous ones, causing errors to accumulate. Also, since the results are only relative measurements rather than absolute, it is still necessary to measure the absolute position and orientation of at least one reference point along the colonoscope to make sense of the results.

The main aim of the project is to produce an image of the configuration of the colonoscope and not to image the patient. If the position of each point along the colonoscope could be measured in 3D space, relative to some external reference co-ordinate system, it would be possible to reconstruct an image of the colonoscope. The position sensing technology used in this method would need to operate in the presence of the patients body. For example, using electromagnetic fields as the basis of the positioning system would not be possible since they are significantly affected by the conductivity and permittivity of the human body. However, static (and quasi-static) magnetic fields penetrate the body but are not significantly affected by

the body, since the permeability of the body is close to that of the air. From a design point of view, magnetic field systems are attractive since they can be generated and detected simply. This is especially true of quasi-static fields, which can be both generated and detected by simple coils of wire. The disadvantage of using magnetic fields is that, unlike electromagnetic waves which basically travel in straight lines, magnetic field lines form closed loops and the position location algorithm needs to accommodate this behaviour.

One method of implementing this idea is to generate magnetic fields under the patient and to detect them inside the colonoscope using magnetic field sensors. These sensors could either be encapsulated in a catheter and inserted into the existing biopsy channel of the colonoscope, or alternatively could be built into the colonoscope during manufacture. The field measurements from these sensors could then be used to calculate the position of each sensor, using a suitable algorithm. Having found the position of each sensor, a curve could be fitted through the points to construct an image of the colonoscope.

It could be argued that this system is technically not an imaging system, since it is locating something inside the object of interest and not the object itself. However, to all intents and purposes it functions as an imaging system, since it presents an image of the colonoscope to the observer, i.e. the clinician.

1.10 Conclusion

From this study it has emerged that a promising option is to develop a 3D positioning system based on magnetic fields. Magnetic fields can be generated and detected using low cost devices and are considered safe. This means that unlike fluoroscopy, an imaging system based on them could be operated continuously throughout the colonoscopy procedure. Also, unlike acoustic and microwave fields, the generators do not need to be in contact with the body or require matching media, making the system convenient to use.

In order to implement the system, a position location algorithm is necessary, which takes the field measurements from the sensors and converts them into sensor positions. In conjunction with this, suitable fields need to be created which enable the algorithm to operate, since careful selection of fields can greatly simplify the algorithm.

1.11 References

[1.1] J.D. Kettner : "Colorectal Cancer: Epidemiolgical and Clinical Issues", Symposium at Royal Society of Medicine, London, Duphar Medical Relations, (1989).

[1.2] J.M.A. Northover and J.D. Kettner : "Bowel Cancer - The Facts", Oxford University Press (1992).

[1.3] T. Muto, H.J.R. Bussey and B.C. Morson : "The evolution of cancer of the colon and rectum", *Cancer*, Vol. 36, pp.2251-2270, 1975.

[1.4] C.B. Williams in "Therapeutic Endoscopy", Ed. by J. Bennett, Chapman and Hall (1981).

[1.5] J.S. Atwater and J.A. Bargen, "Pathogenesis of intestinal polyps", *Gastroenterology*, Vol, 4, pp.395-408, 1945.

[1.6] T.C. Arminski and D.W. McLean : "Incidence and distribution of adenomatous polyps of the colon and rectum based on 1000 autopsy examinations", *Dis. Colon Rect.*, Vol. 7, pp.249-261, 1964.

[1.7] L.E. Hughes : "The incidence of benign and malignant neoplasms of the colon and rectum - a post mortem study", *Aust. N. Z. J. Surg.*, Vol. 38, pp.30-35, 1968.

[1.8] F.E. Silverstein and G.N.J. Tytgat : "Atlas of Gastrointestinal Endoscopy", Gower Medical Publishing, 2nd Edition 1991.

[1.9] J.A. Vennes : "Instrumentation and Disinfection" in "Colonoscopy", Ed. by R.H. Hunt and J.D. Waye, Chapman and Hall (1981).

[1.10] P.B. Cotton and C.B. Williams : "Practical Gastrointestinal Endoscopy", Blackwell Scientific Publications, 3rd Edition.

[1.11] B.R. Parry and S.M. Williams : "Competency and the colonoscopist: a learning curve", *Aust. N.Z. J. Surg.*, Vol. 61, pp. 419-422, 1991

[1.12] C.B. Williams in "Colonoscopy: general aspects, polyps and cancers", Annual of gastrointestinal endoscopy, London: Current Science (1992).

[1.13] C.B. Williams in "Colonoscopy", Practical gastrointestinal endoscopy, 3rd edition (P.B. Cotton, C.B. Williams), Oxford: Blackwell, (1990).

[1.14] R.H. Hunt : "Colonoscopy Intubation Techniques With Fluoroscopy" in "Colonoscopy", Ed by R.H. Hunt and J.D. Waye, Chapman and Hall (1981).

[1.15] W.C. Cirocco and L.C. Rusin : "Documenting the use of fluoroscopy during colonoscopic examination: a prospective study", *Surg. Endosc.*, Vol. 5, pp.200-203, 1991.

[1.16] F. Kee, R.H. Wilson, R. Gilliland, J.M. Sloan, B.J. Rowlands and R.J. Moorehead : "Changing site distribution of colorectal cancer", BMJ, Vol. 305, 18 July 1992.

[1.17] W. Swindell and S. Webb : "X-ray Transmission Computed Tomography" in "The physics of medical imaging", Ed. by S. Webb, IOP Publishing Ltd (1988).

[1.18] P.N.T. Wells : "Ultrasonics in clinical diagnosis", 2nd edition, Churchill Livingstone, (1977).

[1.19] J-Ch. Bolomey and M.S. Hawley : "Noninvasive Control of Hyperthermia" in "Methods of Hyperthermia Control", Ed. M. Gauterie, Springer-Verlag Berlin Heidelberg (1990).

[1.20] C. Pichot, L. Jofre, G. Peronnet, J-Ch. Bolomey : "Active microwave imaging of inhomogeneous bodies", IEEE Trans, AP 33(4), pp 416-425, 1985.

[1.21] A.D. Seagar, D.C. Barber and B.H. Brown : "Theoretical limits to sensitivity and resolution in impedance imaging", Clin. Phys. Physiol. Meas., Vol. 8, Suppl. A, pp13-31, 1987.

[1.22] "X rays of visible light", Innovations, Ed. by T.E. Bell, IEEE Spectrum, p16, Aug 1994.

Chapter 2

Developing A Position Location System That Uses Magnetic Fields

2.1 Using Magnetic Fields For Position Location

The concept of using magnetic fields to image any catheter or similar small tube inserted into the body represents a new challenge and since it is not an established area of research, there are no standard techniques for doing this. The main objective for the research was to devise and evaluate a practical system, suitable for use in clinical practice, that produces a real time image of a colonoscope to the required positional accuracy.

The aim is to create a system that can determine the position of multiple sensors which are placed inside the colonoscope in the form of a catheter. Creating an image of the colonoscope from this information is then the relatively straightforward process of fitting a curve through the points and displaying the result on a computer screen. Although the system must eventually determine the position of multiple sensors, the initial task is to produce a system which can determine the position in three dimensions of a single field sensor. Extending the system into a multiple sensor version is then just a matter of duplicating the sensing channels.

A magnetic field positioning system needs a method of generating the magnetic fields, a method of measuring the fields at the point of interest, and an algorithm which uses a knowledge of the fields to convert the measurements into the sensor position. The orientation of the sensor may be of interest as well, in which case the algorithm may be able to supply this information.

Suitable fields for the position location system are either static or quasi-static magnetic fields since these do not significantly interact with the body. Static magnetic fields are time invariant and are produced by passing a dc current through a loop of wire. Quasi-static fields are produced by passing a low frequency current through a loop of wire, but otherwise behave identically to static fields. What constitutes 'low frequency' depends on the size of the coil, and the interaction with the environment, and in this application is typically up to 100 kHz. The advantage of using quasi-static fields over static fields is that they can be detected by a simple coil. They also avoid the problems of measurement offsets due to the Earth's magnetic field, that would arise if static fields were used.

An early experiment comprised a set of permanent magnets (which produce static fields) and a magnetic compass. To calibrate the system, the field directions were recorded for each point on a 2D grid, for different configurations of the magnets. A search algorithm was then developed which took readings from the compass and in conjunction with the stored data determined the position of the compass in 2D for any orientation of the compass. Remarkably, the method produced an accuracy of a few millimetres over a plane of about 30 cm^2. However, extending this method to solve the problem of colonoscopy is difficult since it is difficult to make a compass which is both capable of remote sensing and is small enough to fit inside the biopsy channel of a colonoscope. Also the method is difficult to extend to 3D and requires re-calibration if the orientation of the equipment relative to the Earth's field is changed. However it proved an interesting introduction to the principles of magnetic field position location.

2.2 Requirements For The Position Location System

The position location system must be designed to operate in the colonoscopy environment. A principle requirement is that the field sensors must be small, typically less than 2 mm diameter, in order that they may fit inside the biopsy channel of a colonoscope. The

positioning system must also produce 3D information about the sensor and must be tolerant to the sensor being pointed in any direction. A further requirement is that the position location algorithm must be fast enough to enable all the sensors in the colonoscope to be located and a real time image of the colonoscope to be displayed in real time. The minimum speed of the algorithm, ignoring the time taken to acquire the data and display the image, is given by:

Points Per Second = (Frames Per Second) * (Number Of Sensors In Catheter) (2.1)

2.3 Techniques For Generating Magnetic Fields

Magnetic fields are generated by passing a current around a closed circuit. The field strength is directly proportional to the magnitude of the current. If the circuit consists of a number of turns of wire, then the field strength is also proportional to the number of turns, since the current passes around the circuit once for each complete loop of wire. The field produced by a given coil can be calculated using a variety of analytic and numeric techniques and these are discussed later.

2.4 Techniques For Detecting Magnetic Fields

One method for detecting quasi-magnetic fields is by the use of a coil. A single coil measures only the field component which is resolved along its axis, and to measure the magnitude and direction of the field, three orthogonal coils are necessary. The sensitivity of a sensor coil is a function of the size of the coil, the number of turns and the frequency of the field. Manufacturing a sensor comprising a single coil which is small enough (2 mm diameter) for our application is straightforward. However it would be more difficult to produce a three axis sensor which is small enough.

An alternative type of sensor is a Hall effect probe. These typically have a sensitivity of 0.01 mV per μT, and a minimum detectable flux density of about 100 μT. For the strength of fields that can be conveniently generated over the required volume, this sensitivity is inadequate. Size, thermal drift and noise are further problems with using these devices for this application.

Fluxgate magnetometers [2.1] are more sensitive than Hall probes and have a resolution of around 1 μT. They consist of a toroidal core wound with two sets of windings. The primary winding is pulsed with a current which drives the core into saturation, and the secondary is wound in such a way that the fields cancel, giving no net pick up in the secondary winding. Any external field unbalances the cancellation effect and results in a net positive or negative signal in the secondary. Fluxgate magnetometers are usually at least several centimetres in diameter, so they are unsuitable for use inside the patient.

SQUID detectors, which require super cooling, are highly sensitive and are typically used to detect the very weak magnetic fields produced by signals in the brain. They are capable of detecting fields as small as a few tens of fT. Although it is inappropriate for use inside the body, a SQUID detector could be used outside the patient if the situation was reversed and the field was generated inside the colonoscope.

2.5 Existing Point Location Technologies

Systems utilising magnetic fields for short range position measurement are commercially available [2.2, 2.3, 2.4, 2.5, 2.6]. They are used for a variety of applications including tracking vehicles, locating buried miners and locating the head of a pilot. They are also found in computer user interface applications such as the 3D mouse.

The majority of the commercial systems that are available [2.4, 2.5] involve field sensors which comprise three coils wound orthogonally. This enables both the magnitude and direction of the field to be measured at a point and enables the sensor position to be calculated fairly simply. Other systems use a two coil sensor, but give only 2D positional information and require the sensor to operate in a plane. Systems having just a single coil sensor have severe limitations, for example requiring mechanical movement of the sensor during the measurement process [2.2] or producing only 2D information [2.6].

Polhemus [2.4, 2.5, 2.7] offer a wide range of systems which are used to determine positions within a total range of one metre. These systems consist of a field generating module, a field sensing module and a control unit, thereby determining the x, y, z co-ordinates and θ, ϕ, ψ orientation of the sensor relative to the generator to an accuracy of a few millimetres. Both the generator and the sensor modules comprise three coils. The location principle operates by steering the dipole field, for example by varying the drive current to each of the three generator coils, until the axis of the field is directed toward the sensor. Once this is achieved, the distance from the generator module to the sensor module can be calculated from the magnetic dipole field equation [2.8]. The angles are calculated by solving a set of five 3D rotations which transform the fields produced by the generator module into the field measured by the sensor module. To help solve these transformations, a system known as *nutation* is used which involves the dipole direction (called the pointing vector) being made to circle around the sensor. This also helps to track the sensor when it is moved. A block diagram of this system is shown in figure 2.1. The angles A and B describe the direction of the pointing vector, and θ, ϕ and ψ describe the orientation of the sensor.

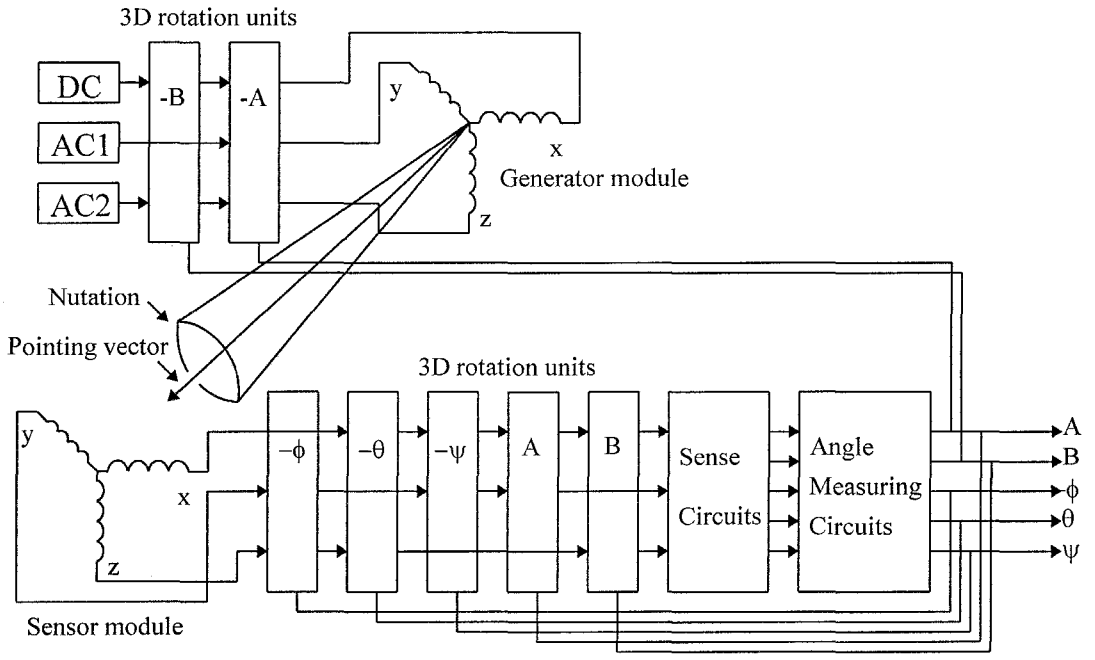


Fig 2.1: Polhemus Position Location System

In principle, the accuracy and range of this system is suitable for medical applications, but the sensor module is inherently too large. Also, three coils are needed for each sensor and since around 16 sensors are needed along the length of the colonoscope, a total of 48 coils are needed, connected by 96 wires. Unfortunately, it was not possible to produce even a single sensor small enough to fit inside the biopsy channel. The sensitivity of the sensor decreases with size, and so miniaturising the sensor results in a poor signal to noise ratio. Also the metal present in the colonoscope will cause the coils to appear non-orthogonal, causing significant measurement errors. In fact the system was found to become unstable in the presence of any metal objects, even distant from the field generator and sensor modules.

2.6 Developing A System For Single Coil Sensors

Given the problems in producing three coil sensors which are small enough to fit inside the colonoscope biopsy channel, it seemed sensible to consider using single coil sensors instead. This type of sensor is capable of considerable miniaturisation but produces less information and therefore cannot be used with existing position location systems. However, given the

practical advantages it seemed worthwhile investigating the possibility of developing a new magnetic field position location system.

2.7 Conclusion

From this examination of magnetic field positioning systems, it seems that a system which is capable of locating miniature sensors could be used to develop an endoscope imaging system. A disadvantage of this method is that it requires a catheter to be inserted into the colonoscope while the system is being used. However, under normal circumstances the biopsy channel is not required during intubation. The colon wall is examined in detail as the colonoscope is inserted, and the procedures which require the biopsy channel, such as polyp removal, are performed as the colonoscope is removed. This prevents blood from obscuring the view of the colon. Once the colonoscope is fully inserted, the imaging system becomes less important and the catheter can be removed, releasing the biopsy channel for use by other instruments. The benefit of using the biopsy channel for the position sensors is that the system can be used with existing colonoscopes without modification. It could subsequently be made more convenient to use by integrating the sensors into the colonoscope during the manufacture of the instrument.

Having excluded other types of magnetic sensors on the bases of size, sensitivity and practicality, it seems clear that the most practical sensor to use is a single coil, miniaturised to fit inside the biopsy channel. This requires an entirely new position algorithm and system. Also, since it is necessary to locate the entire path of the colonoscope, not just the tip, multiple sensors together with image construction software are also needed.

2.8 References

[2.1] R. Noble : "Fluxgate magnetometery", Electronics World and Wireless World, pp 726-732, Vol. 97 No. 1666, Sept 1991.

[2.2] Patent : H.R. Weed and R.M. Engira : "Remote position and orientation detection employing magnetic flux linkage", US Pat. No. 4317078, 1982.

[2.3] Patent : H.P. Kalmus : "Direction Indicator", US Pat. No. 3121228, 1964.

[2.4] Patent : J. Kuipers : "Object tracking and orientation determination means, system and process", US Pat. No. 3868565, 1975.

[2.5] Patent : J. Kuipers : "Tracking and determining orientation of object using coordinate transformation means, system and process", US Pat. No. 3983474, 1976.

[2.6] Patent : "Locating Process For Localising Unknown Receiver Or Transmitter Positions", Pat. Pub. No. WO 88/09515, International Application No. PCT/AT88/00036, 1988.

[2.7] J.B. Kuipers : "SPASYN - An electromagnetic relative position and orientation tracking system", IEEE Trans. on Inst. and Meas., Vol. IM-29 No. 4, pp 462-466, Dec 1980.

[2.8] M. Zahn : "Electromagnetic Field Theory", pp 344-346, J. Wiley and Sons (1979).

Chapter 3

Theoretical Development Of The Point Location Algorithm

3.1 Principles Of A Point Location Algorithm For A Single Coil Sensor

The purpose of the point location algorithm is to determine the position and orientation of a single coil sensor within one or more low frequency magnetic field distributions. The sensor coil produces a signal whose amplitude is proportional to the magnitude of the component of field in the sensor. Also the phase of the signal, with respect to the current in the field generator coil, indicates the direction of the field through the sensor.

The position and orientation of a single coil sensor can be uniquely described by three position parameters x, y, z and two orientation parameters θ and ϕ. A rotation ψ about the axis of the coil does not produced a change in the signal from the sensor and therefore does not constitute an additional parameter. The parameters x, y, z, θ, ϕ are measured with respect to a fixed reference co-ordinate frame.

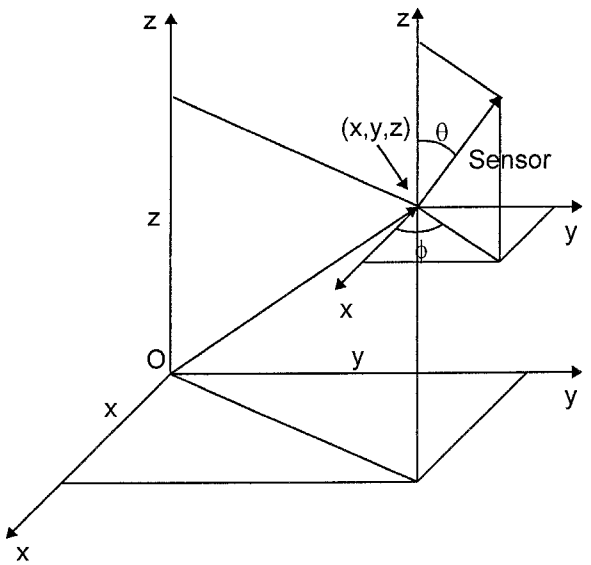


Fig 3.1: Co-ordinate system showing the five parameters which describe the position and orientation of a single coil sensor (x, y, z, θ, ϕ)

The voltage induced in the sensor can be written as a function of the five position variables:

$$v = f(x, y, z, \theta, \phi) \tag{3.1}$$

Since there are five parameters which must be found in order to determine the position and orientation of the sensor, it follows that a minimum of five equations are necessary to provide enough constraints to determine these parameters. These equations may be written:

$$v_i = f_i(x, y, z, \theta, \phi) \quad (i = 1 \text{ to } 5) \tag{3.2}$$

where i is the number of the field. Clearly the five equations must be different in some way and this requires different fields to be created. For example, one way of doing this, which works provided the field generators do not produce a uniform field, is to place the field generators at different locations.

To proceed, a suitable type of field generator must be chosen. In general, the magnetic field produced by a loop of wire of arbitrary shape cannot be represented by a simple analytic equation and needs to be stored as a large table of numbers. It is difficult to develop an efficient point location algorithm for such a field, since it will involve a multi-dimensional (in this case 5D) search through the table of field values. In practice, this is not feasible due to the number of results which must be stored and searched. For example, scanning a 1 m^3 volume with a 10 mm resolution and 10 degree angle increments, would result in 1.3×10^9 points. It is clearly impractical to scan, store and search such a large amount of data.

3.2 The Magnetic Dipole

Fortunately, certain types of field generator produce fields which can be represented by simple analytical equations. One such generator is the magnetic dipole, which consists of a

small loop of current. 'Small' means that the size of the coil is small compared with the distance between the coil and the point in the field of interest.

The magnetic field at a point P, produced by a square dipole at the origin, is given by equation 3.3 [3.1, 3.2]. The shape of the coil affects only the multiplying constant k_G and does not affect the shape of the field.

$$\mathbf{B} = \frac{k_G}{R^3}\left[2\hat{\mathbf{a}}_R \cos\alpha + \hat{\mathbf{a}}_\alpha \sin\alpha\right] \tag{3.3}$$

where,

$\mathbf{B}$ = Magnetic flux density (Tesla)

R = Distance from dipole centre to P (metres)

α = Angle from coil axis to P (radians)

$$k_G = \frac{\mu_0 NIb^2}{4\pi}$$

$\mu_0 = 4\pi \cdot 10^{-7}$

N = number of turns

I = coil current (Amps)

b = side of coil (metres)

$\hat{\mathbf{a}}_R$ = unit vector in direction of R.

$\hat{\mathbf{a}}_\alpha$ = unit vector in direction of α.

The field magnitude is given by:

$$|\mathbf{B}| = \frac{k_G}{R^3}\left[\sqrt{3\cos^2\alpha + 1}\right] \tag{3.4}$$

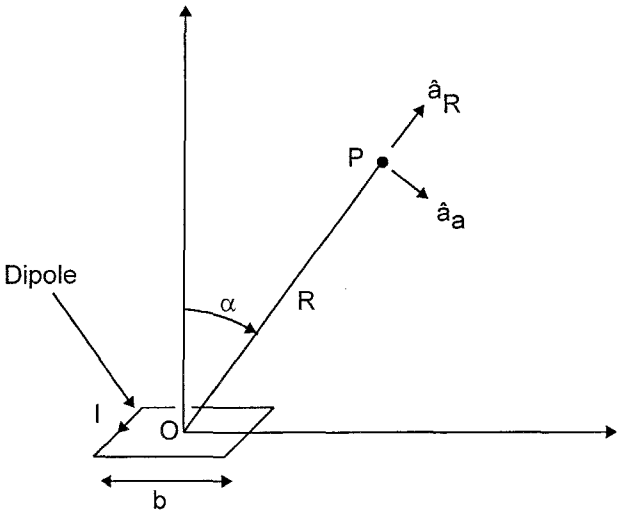


Fig 3.2: Co-ordinate system for the dipole field equation

The term *dipole* is misleading since it implies 'two poles' when it fact the field is produced by a current loop. The term may stem from an early belief that separate North and South poles exist. If a pole did exist separately then the divergence at that point would be non-zero, which disagrees with the Maxwell equation $\nabla \cdot \mathbf{B} = 0$. However the term *dipole* is still commonly used, and therefore is used here.

A further complication is that some early text books derive the magnetic dipole field equation from the electric dipole case, replacing the point charges with North and South poles. (They also talk about 'effective pole positions' which are outside the coil, since if the poles are inside the coil, and the coil is a single turn of wire, then the poles must be at the same position. Using their analogy with the electric field case, this would mean that the coil will produce no field, which is clearly not the case). Ironically, this approach produces the correct answer (except for the multiplying constant) and the reason for this is interesting. At distances which are large compared with the size of the dipole, the curl and divergence of both the magnetic field and electric field are zero, and it follows from this that the field equations for the two different sources will have the same form [3.2]. Close to the dipoles though, the field shapes are quite different.

A coil which produces a field which exactly obeys equation 3.3 is referred to henceforth as an *ideal dipole*. However, to produce this field the coil must be of zero size and consequently will produce no field. In practice a coil must have a finite size in order to produce a field and therefore produces only an approximation to the dipole field. Such a generator is referred to as a *practical dipole*. In the limit where R >> b, the practical dipole field tends to that of an ideal dipole field.

The field error in using practical dipoles, together with the positional error which arises from it is calculated later. In any practical measurement system making the dipole coil too large increases the error in the field approximation, whereas making the coil too small increases measurement errors due to a reduced signal to noise ratio in the measurement of the field. Thus for a given application, there is an optimum generator coil size of coil.

3.3 Acquiring The Measurement Data

In order to solve equation 3.2, five different dipole fields need to be generated. Clearly the five field generator coils cannot be coincident, since then they will produce identical fields. Nor can the field shape be altered by changing the drive to the coils. The only way to create different fields is by having different positions and/or orientations for each generator coil.

The sensor voltage must be measured for each of the five fields. One way of doing this is to energise each generator in turn and for each generator, measure the induced voltage in the sensor. This is the sequential method. Alternatively, all the generators could be energised simultaneously using different drive frequencies or other signal coding. This method also requires some form of filtering or decoding of the sensor signal. Although the second method may acquire the measurement data more quickly, it requires more sophisticated hardware and a higher bandwidth. For this reason, the first method has been used exclusively.

3.4 Fields Produced By The Generator Coils

The first method considered to obtain the fields was to place all the generators on the z=0 plane with their axes pointing in the +z direction, but with their centres at different x and y positions. The z=0 plane is essentially an arbitrary reference plane to which all position measurements are made. Figure 3.2 shows a plan view of the five generators:

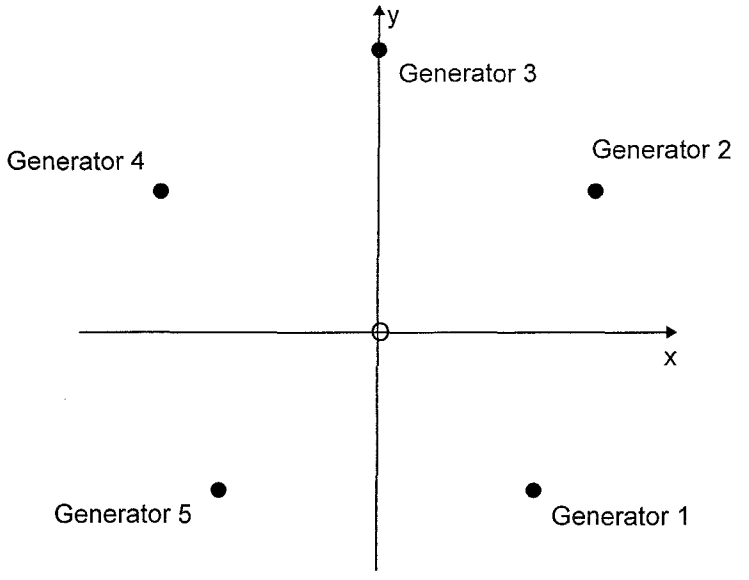


Fig 3.3: Plan view of the five field generator coils

The field produced by each generator coil can be derived from equation 3.3. For a dipole situated at the origin, with its axis pointing in the +z direction, the Cartesian form of the field can be derived as follows:

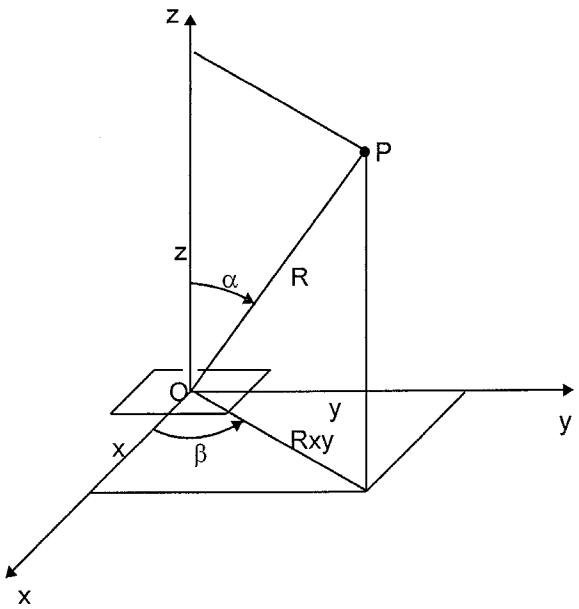


Fig 3.4: Deriving the Cartesian form of field for a dipole at the origin.
The diagram shows a dipole coil at the origin, pointing in the +z direction.

$$R = \sqrt{x^2 + y^2 + z^2} \tag{3.5}$$

$$R_{xy} = \sqrt{x^2 + y^2} \tag{3.6}$$

$$B_R = \frac{k_G}{R^3}(2\cos\alpha) \tag{3.7}$$

$$B_\alpha = \frac{k_G}{R^3}(\sin\alpha) \tag{3.8}$$

$$\cos\alpha = \frac{z}{R} \tag{3.9}$$

$$\sin\alpha = \frac{R_{xy}}{R} \tag{3.10}$$

$$\cos\beta = \frac{x}{R_{xy}} \tag{3.11}$$

$$\sin\beta = \frac{y}{R_{xy}} \tag{3.12}$$

Resolving B_R and B_α:

$$B_{xy} = B_R \sin\alpha + B_\alpha \cos\alpha \tag{3.13}$$

$$= \frac{k_G}{R^3}\left[2\cos\alpha\sin\alpha + \sin\alpha\cos\alpha\right] \tag{3.14}$$

$$= \frac{k_G}{R^3}\left[3\cos\alpha\sin\alpha\right] \tag{3.15}$$

$$= \frac{k_G}{R^3}\left[3\frac{zR_{xy}}{R^2}\right] \tag{3.16}$$

$$B_x = B_{xy}\cos\beta \tag{3.17}$$

$$= \frac{k_G}{R^5}(3xz) \tag{3.18}$$

$$B_y = B_{xy}\sin\beta \tag{3.19}$$

$$= \frac{k_G}{R^5}(3yz) \tag{3.20}$$

$$B_z = B_R\cos\alpha - B_\alpha\sin\alpha \tag{3.21}$$

$$= \frac{k_G}{R^3}\left[2\cos^2\alpha - \sin^2\alpha\right] \tag{3.22}$$

$$= \frac{k_G}{R^5}\left[2z^2 - x^2 - y^2\right] \tag{3.23}$$

Combining the field components gives:

$$\mathbf{B} = \frac{k_G}{R^5}\left[3z\left(\hat{\mathbf{a}}_x x + \hat{\mathbf{a}}_y y\right) + \hat{\mathbf{a}}_z\left(2z^2 - x^2 - y^2\right)\right] \tag{3.24}$$

To produce the field equations for each generator, it is necessary to translate this equation from the origin to the position of each generator. If the co-ordinates of each generator are $(x_{Gi}, y_{Gi}, 0)$, where i is the number of the generator, then we can find the field produced by generator i by replacing each x term with $x - x_{Gn}$, and each y term with $y - y_{Gn}$.

$$\mathbf{B}_{Gi} = \frac{k_G}{R^5}\left[3z\left(\hat{\mathbf{a}}_x(x - x_{Gi}) + \hat{\mathbf{a}}_y(y - y_{Gi})\right) + \hat{\mathbf{a}}_z\left(2z^2 - (x - x_{Gi})^2 - (y - y_{Gi})^2\right)\right] \quad (i = 1..5) \quad (3.25)$$

3.5 Emf Induced In The Sensor

The emf induced in the single coil sensor when placed in a field is proportional to the resolved component of the field along the axis of the sensor. This component may be found by computing the dot product between the field vector at the sensor and the unit vector $\hat{\mathbf{a}}_S$ which lies along the axis of the sensor:

$$V_S = k_S(\mathbf{B} \cdot \hat{\mathbf{a}}_S) \quad (3.26)$$

where:

V_S = emf induced in the sensor (Volts)

k_S = sensitivity of the sensor (Volts per Tesla)

$\hat{\mathbf{a}}_S$ = unit vector in the direction of the sensor

$\hat{\mathbf{a}}_S$ may also be represented in terms of two angles, θ and ϕ, as shown in figure 3.5.

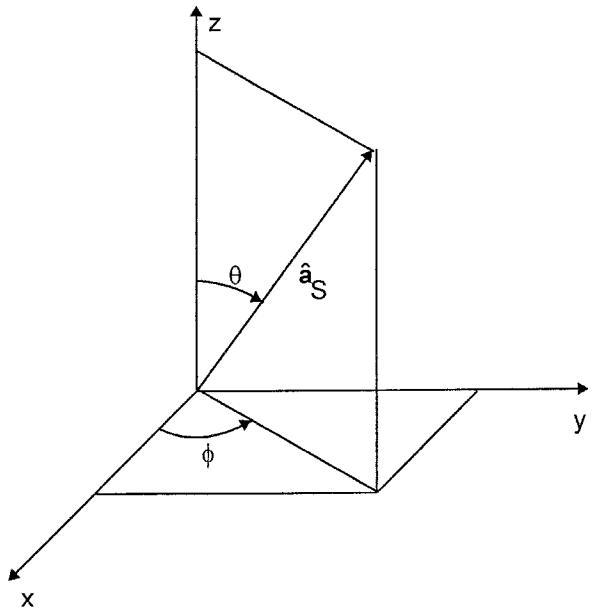


Fig 3.5: Angles defining orientation of the sensor

From figure 3.5, the components of $\hat{\mathbf{a}}_S$ can be written:

$$S_x = \sin\theta\cos\phi \tag{3.27}$$

$$S_y = \sin\theta\sin\phi \tag{3.28}$$

$$S_z = \cos\theta \tag{3.29}$$

Computing the dot product gives:

$$V_S = k_S\left[B_x \sin\theta\cos\phi + B_y \sin\theta\sin\phi + B_z \cos\theta\right] \tag{3.30}$$

From equations 3.25 and 3.30, the induced emf in the sensor for a given position and orientation of the sensor is given by:

$$V_{Si} = \frac{k_G k_S}{R^5}\left[3z\sin\theta\left((x-x_{Gi})\cos\phi + (y-y_{Gi})\sin\phi\right) + \cos\theta\left(2z^2 - (x-x_{Gi})^2 - (y-y_{Gi})^2\right)\right]$$

(i=1..5) (3.31)

This equation is in the same form as equation 3.2. The next step is to develop a method of solving the five simultaneous non-linear equations.

3.6 Solving Non-Linear Simultaneous Equations

Equation 3.31 represents a set of non-linear simultaneous equations which must be solved to find x, y, z, θ and ϕ. In general, the problem of solving simultaneous non-linear equations is a difficult problem [3.3, 3.4, 3.5]. A principle difficulty is preventing the solution from getting trapped in local minima rather than finding the global minima.

One approach, called Newton's method [3.4], attempts to solve the case where the number of equations (m) is the same as the number of unknowns (n). It can be described as follows:

$$\mathbf{J}_k \mathbf{p}_k = -\mathbf{f}_k \tag{3.32}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k \tag{3.33}$$

where,

J = Jacobian matrix, defined as:

$$\mathbf{J} = \begin{bmatrix} \partial f_1/\partial x_1 & \cdots & \partial f_1/\partial x_n \\ . & & . \\ \partial f_m/\partial x_1 & \cdots & \partial f_m/\partial x_n \end{bmatrix} \tag{3.34}$$

p = vector

f = $f_1, f_2 \ldots f_m$ = set of functions to solve, rearranged to equal zero at solution.

x = $x_1, x_2 \ldots x_n$ = set of variables to solve

k = number of iterations

The method is used as follows:

1. Estimate **x**.
2. Calculate **J** and **f**.
3. Solve m simultaneous linear equations in n unknowns to find **p**.
4. Calculate the new estimate of **x**.
5. Check for convergence. Go to step (2) if not converged.

Another approach, which copes with the case where n <> m is called the Gauss-Newton method [3.3, 3.5]. It attempts to minimise the function:

$$F(x) = \sum_{i=1}^{m} f_i^2(x) \tag{3.35}$$

The Gauss-Newton method can be described by the following equations:

$$\mathbf{J}_k^T \mathbf{J}_k \mathbf{p}_k = -\mathbf{J}_k^T \mathbf{f}_k \tag{3.36}$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k \tag{3.37}$$

where,

J = Jacobian matrix, defined as:

$$\mathbf{J} = \begin{bmatrix} \partial f_1/\partial x_1 & \dots & \partial f_1/\partial x_n \\ . & & . \\ \partial f_m/\partial x_1 & \dots & \partial f_m/\partial x_n \end{bmatrix} \tag{3.38}$$

$\mathbf{J}^T$ = matrix transpose of **J**

p = vector

$\mathbf{f} = f_1, f_2 \dots f_m$ = set of functions to solve, rearranged to equal zero at solution.

$\mathbf{x} = x_1, x_2 \dots x_n$ = set of variables to solve

k = number of iterations

The method is used as follows:

1. Estimate **x**.
2. Calculate **J**, $\mathbf{J}^T$ and **f**.
3. Solve m simultaneous linear equations in n unknowns to find **p**.
4. Calculate new estimate of **x**.
5. Check for convergence. Go to step (2) if not converged.

The first attempt to solve equation 3.31 used Newton's method with 5 equations and 5 unknowns. To avoid possible errors or bugs in the measurement hardware from causing problems, initial tests were performed on the algorithm using simulated data rather than actual measurement data. The simulated data were produced by generating a series of random sensor positions and then calculating the sets of sensor voltages that would be obtained at each of these positions using equation 3.31. The algorithm was then applied to each set of data, and the positional results obtained compared with the original positions.

The algorithm performed erratically, either converging onto the correct solution, or onto a totally different position. In an attempt to improve the situation, various different generator positions and orientations were tried, but this did not improve the reliability of the method. The problem was that different sensor positions and orientations could give rise to almost identical sets of measurement data. That is, there wasn't a one-to-one mapping between sensor position and measurement data. Thus the solution could easily converge onto an incorrect alternative sensor position.

Therefore, although Newton's method is known to be an imperfect technique [3.3], the problem of incorrect solutions was clearly not due to shortcomings in the method, but rather was due to the choice of fields and field equations. Since the measurement data did not always represent the sensor's position uniquely, an attempt was made to improve the situation by increasing the number of measurements. This was achieved by increasing the number of field generators to 8, 16 and 32. Since this means that Jacobian matrix is no longer square (that is m<>n), Newton's method was replaced by the Gauss-Newton method (equation 3.36).

The new algorithm was tested using the same simulated data that was used previously, but performed no better, producing more incorrect solutions rather than less. Rather that pursue other standard solution methods, the form of the field equations was re-examined.

3.7 Strategy Of Combining The Generator Coils

So far the equation 3.31 has proved difficult to solve. The general form of these equations is fixed, but it is possible to combine the generators so that the equations become easier to solve. This can be done by grouping the generator coils into orthogonal coil assemblies, where each assembly consists of three generator coils. For reasons which will become evident later, three such assemblies were used, giving a total of nine generator coils.

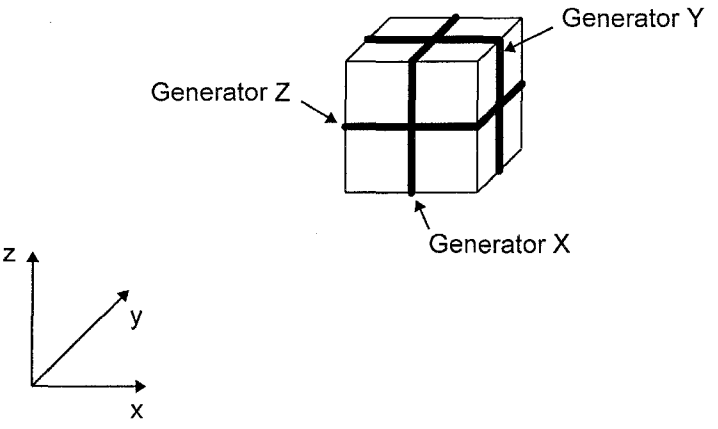


Fig 3.6: Orthogonal generator coil

At this stage it is conceptually advantageous to imagine that the rôles of the generator and the sensor coils are reversed such that the sensor is producing the field and the generator is sensing the field. This is helpful because it is equivalent to having a single field whose amplitude and direction is measured at different points, rather than multiple fields where resolved component of the fields are measured at a single point. The equations which describe the induced sensor voltage for a given generator current remain the same since this

relationship depends on the mutual inductance between two coils, and is independent of which coil is the generator and which is the sensor.

Figure 3.7 shows the sensor and a single orthogonal generator coil assembly. α is taken to be the smaller angle between the sensor direction $\hat{\mathbf{a}}_S$ and the radial vector **R**.

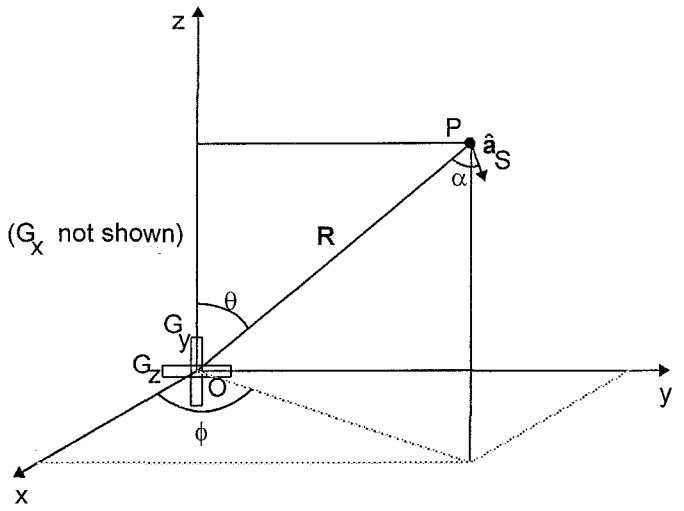


Fig 3.7: Generator and sensor, showing angle α.

The field at O due to $\hat{\mathbf{a}}_S$ (remembering that the sensor and generator are reversed) is given by:

$$\mathbf{B} \propto \frac{1}{R^3}\left[2\hat{\mathbf{a}}_R \cos\alpha + \hat{\mathbf{a}}_\alpha \sin\alpha\right] \tag{3.39}$$

The three orthogonal coils at O will produce a voltage which is proportional to the resolved components of this field along their axes. Since the coils are orthogonal, the three voltages can be combined using Pythagoras to produce a number which is proportional to the magnitude of the field.

$$|\mathbf{B}| \propto \sqrt{V_{Gx}{}^2 + V_{Gy}{}^2 + V_{Gz}{}^2} \tag{3.40}$$

From equation 3.39, the field magnitude can also be written as:

$$|\mathbf{B}| \propto \frac{1}{R^3}\left[\sqrt{3\cos^2\alpha+1}\right] \tag{3.41}$$

From equations 3.40 and 3.41:

$$\sqrt{V_{Gx}{}^2+V_{Gy}{}^2+V_{Gz}{}^2} \propto \frac{1}{R^3}\left[\sqrt{3\cos^2\alpha+1}\right] \tag{3.42}$$

Reverting to the actual situation, $_{Gx}$, $_{Gy}$ and $_{Gz}$ represent the induced voltage in the sensor when each of the generator coils are energised. Also, the constant of proportionality is given by $k_S k_G$. The only unknowns in the equations are therefore R and α. Since the $\sqrt{3\cos^2\alpha+1}$ term can only take values of between 1 and 2, we can calculate a fairly good estimate for R from this equation by setting α to an arbitrary value.

In order to locate the sensor position, three values of R are required. Therefore, three generator coil assemblies G_i are used, each placed at different positions on the z=0 plane, as before. This gives three different radial distances, R_i, each from different points on the reference plane.

For generator i, we can write:

$$\frac{R_i{}^3}{\sqrt{3\cos^2\alpha_i+1}} = \frac{k_S k_G}{\sqrt{V_{Gxi}{}^2+V_{Gyi}{}^2+V_{Gzi}{}^2}} \tag{3.43}$$

The three values of R_i can be intersected to locate the sensor:

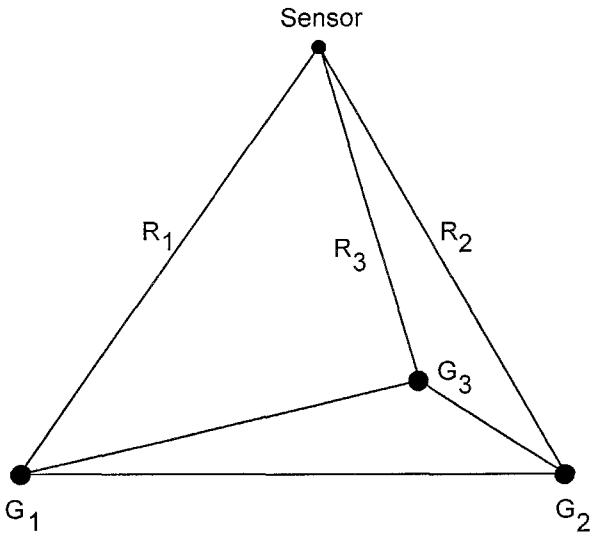


Fig 3.8: Intersecting R_1, R_2 and R_3 to locate the sensor

The values of R_i given by equation 3.43, and hence the estimated sensor position, will not be accurate since α_i is not yet known. However they do constrain the positional solution, and this helps in developing a stable iterative algorithm for locating the precise position.

3.8 Possible Point Location Algorithm 'A'

The first algorithm to use the orthogonal generators, referred to as 'Point Location Algorithm A', is a combination of equation 3.43 and the Gauss-Newton method. The algorithm operates as follows. Initially equation 3.43 is used to calculate an estimate of the distances between the generators and the sensor, by setting α_i to arbitrary values. Having found these distances, they are intersected to find an estimate of the x, y, z position of the sensor. The least squares method is then used to determine θ and ϕ using the estimates of x, y and z. All nine equations are used in this process, even though only two variables are being determined. Although this can theoretically lead to problems of over specification, it avoids having to decide on some criterion for the best pair of equations to use. From these estimates for θ and ϕ, values of α_n are calculated using trigonometry, and in turn are used to calculate better estimates of R_n. Thus the algorithm consists of a two stage process, first finding the sensor

position and then the orientation, that is repeated until the required degree of accuracy is obtained. A flow chart of the process is showed in figure 3.9.

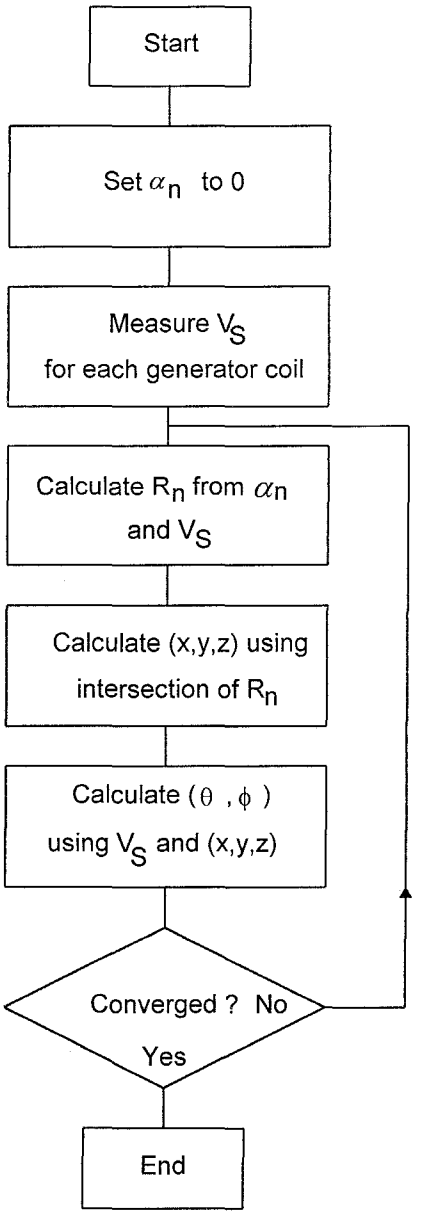


Fig 3.9: Flow chart of algorithm A

This algorithm functions well, giving no erroneous results. In fact, the first medical trials were carried out using this algorithm. The only problem with was that, with the computing power available, it was not fast enough to produce a real time image. On a 486DX2-66 PC, about 24 points per second can be located, resulting in about one and a half image of the catheter being displayed per second, even ignoring the time taken to display the image. This is

not sufficient to give the colonoscopist good feedback, and 8 to 10 frames per second was considered to be more suitable.

3.9 Preferred Point Location Algorithm 'B'

By far the slowest part of algorithm A is the optimisation process used for calculating θ and ϕ, and so an alternative faster method was sought for finding these. The rest of the algorithm remains almost the same. As before, it is helpful to temporarily imagine that the sensor and generator coils are reversed. Thus we can say that there is one generator and nine sensors. (Note that the labelling for S and G_i are not reversed in the diagrams). The strategy is to use the field measurements to calculate the direction of the field at the position of each of the orthogonal generators, which would occur if the sensor and generators were actually reversed. We can then use these, plus a knowledge of the field, to calculate the orientation of the field source (which in this case is the sensor) and hence find α_i. This estimate for α_i can then be used to calculate a better estimate for R_i.

Let the angle between the field at generator G and the radial vector R be ψ:

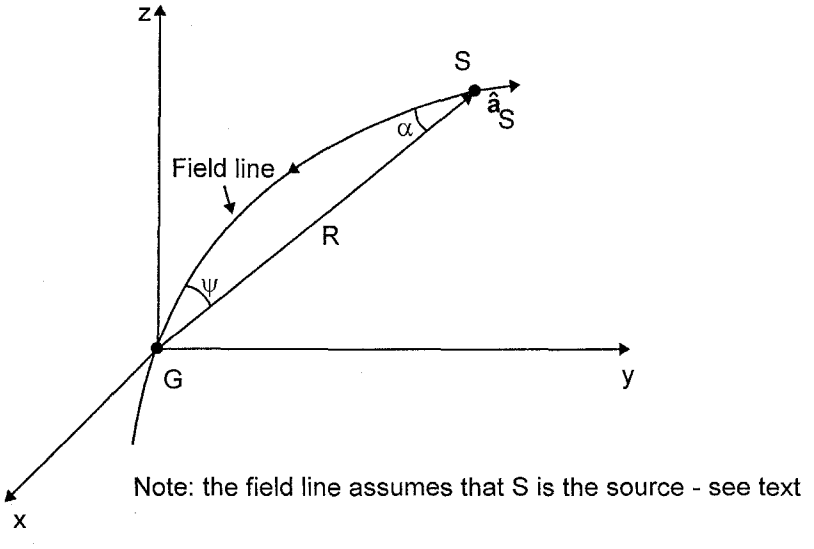


Fig 3.10

The angle ψ can be calculated from the field measurements using the dot product, as follows. The three component vector $\mathbf{V}$ represents the three voltages measured at G, and $\mathbf{R}$ is the radial vector from G to S.

$$\mathbf{V}\cdot\mathbf{R} = |\mathbf{V}||\mathbf{R}|\cos\psi \tag{3.44}$$

$$\cos\psi = \left[\frac{\mathbf{V}\cdot\mathbf{R}}{|\mathbf{V}||\mathbf{R}|}\right] \tag{3.45}$$

or for the general case, where i is the number of the generator coil assembly:

$$\cos\psi_i = \left[\frac{\mathbf{V}_i\cdot\mathbf{R}_i}{|\mathbf{V}_i||\mathbf{R}_i|}\right] \tag{3.46}$$

Having found ψ we need to find α. In figure 3.11, ψ is the known angle and α is the angle we need to find.

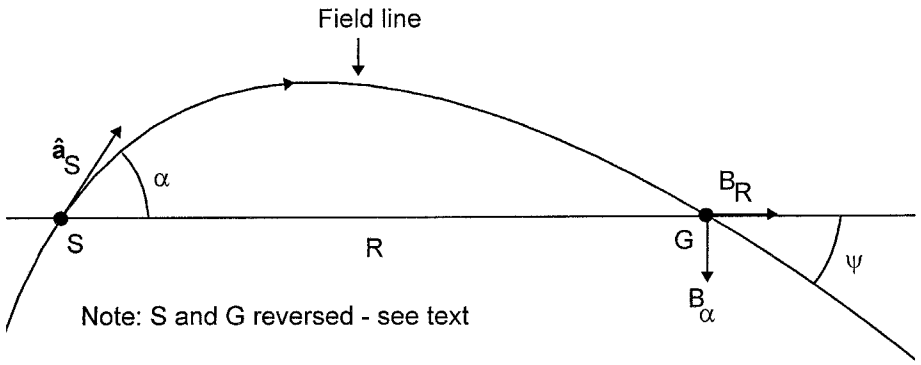


Fig 3.11

Since the generator is a magnetic dipole, the field at the S is given by:

$$B_R = \left(\frac{k_S}{R^3}\right)2\cos\alpha \tag{3.47}$$

$$B_\alpha = \left(\frac{k_S}{R^3}\right)\sin\alpha \tag{3.48}$$

Therefore, the angle of the field at the sensor is given by:

$$\tan\psi = \frac{-B_\alpha}{B_R} = -\frac{1}{2}\tan\alpha \tag{3.49}$$

and so α can be found from ψ using:

$$\tan\alpha = -2\tan\psi \tag{3.50}$$

or in the general case:

$$\tan\alpha_i = -2\tan\psi_i \tag{3.51}$$

Therefore α_i is found by using equation 3.46 to find ψ_i and then using equation 3.51 to find α_i. This is a computationally less expensive method for finding α_i than was used in algorithm A.

The revised algorithm is shown in figure 3.12.

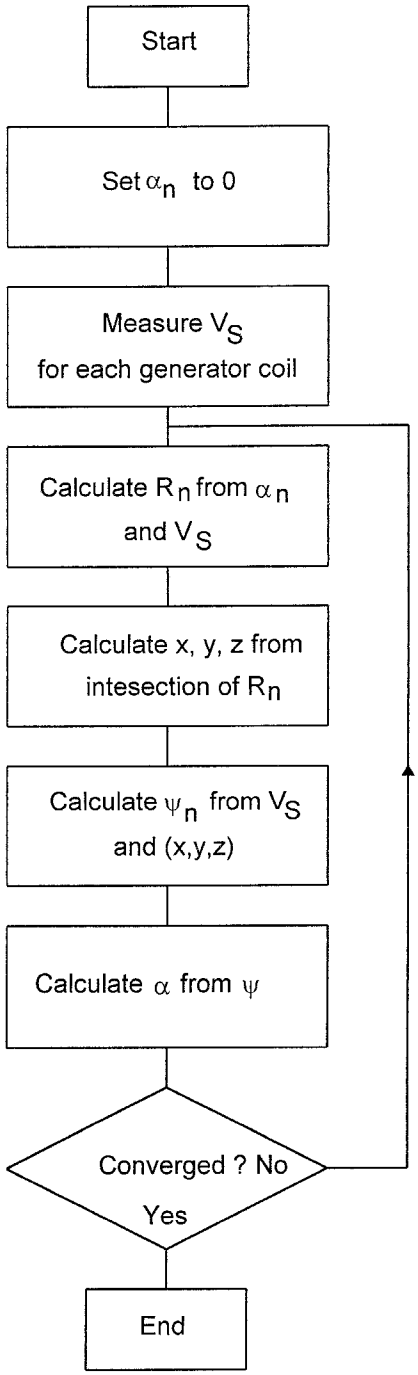


Fig 3.12: Flow chart of algorithm B

Unlike algorithm A, this algorithm does not produce θ and ϕ, and so they need to be calculated after the iterative procedure has been completed. This is done by first finding $\hat{\mathbf{s}}_S$, the unit vector in the direction of the sensor, and then calculating θ and ϕ. To find $\hat{\mathbf{s}}_S$, we

can set up three linear equations for the dot products between $\hat{\mathbf{a}}_S$ and the radial vectors $\mathbf{R}_i$, and then solve them to find the components of $\hat{\mathbf{a}}_S$:

$$\mathbf{R}_i \cdot \hat{\mathbf{a}}_S = |\mathbf{R}_i| \cos\alpha_i \qquad (i = 1 \text{ to } 3) \tag{3.52}$$

The components of **S** are the only unknowns and can be found by solving the simultaneous equations. The sensor orientations θ and ϕ are then found using a four quadrant arctan function:

$$\theta = \arctan\left[\frac{\sqrt{S_x^{\,2} + S_y^{\,2}}}{S_z}\right] \tag{3.53}$$

$$\phi = \arctan\left[\frac{S_y}{S_x}\right] \tag{3.54}$$

Thus the algorithm can produce the five parameters x, y, z, θ and ϕ. The algorithm is more than ten times faster than algorithm A, producing about 350 points per second on a 486DX2-66. This is sufficiently fast to meet the real time goals.

3.10 Conclusion

Two different algorithms have been developed which are able to reliably locate single coil sensors. Both algorithms require three field generator assemblies, each comprising three orthogonal generator coils, and use nine field measurements to determine the sensor position. Preliminary tests have been carried out for immunity to noise and other measurement errors, and both algorithms behave well.

3.11 References

[3.1] M. Zahn : "Electromagnetic Field Theory", pp 344-346, J. Wiley and Sons (1979).

[3.2] R.P. Feynman, R.B. Leighton and M. Sands : "The Feynman lectures on physics", Addison-Wesley Publishing Company (1964).

[3.3] L.E. Scales : "Introduction to Non-Linear Optimistation", MacMillan Publishers Ltd (1985).

[3.4] W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling : "Numerical Recipes In Pascal", Cambridge University Press (1989).

[3.5] L.C.W. Dixon : "Non-Linear Optimisation", The English Universities Press Ltd (1973).

Chapter 4

Effect of Realistic Magnetic Field Distributions On The Position Algorithms

4.1 Magnetic Field Produced By A Square Coil

In Chapter 3, two position algorithms were developed and tested, assuming that the field generators produce ideal dipole fields. However, practical field generator coils do not produce ideal dipole fields and so the algorithm needs to be tested for the non ideal fields that they do produce.

To calculate the fields produced by practical generator coils, Biot-Savarts law [4.1] is used. This law calculates the fields component at a point P due to a current element $I\delta l$ and is defined as follows:

$$\delta B = \frac{\mu_0 I \delta l \sin\beta}{4\pi r^2} \tag{4.1}$$

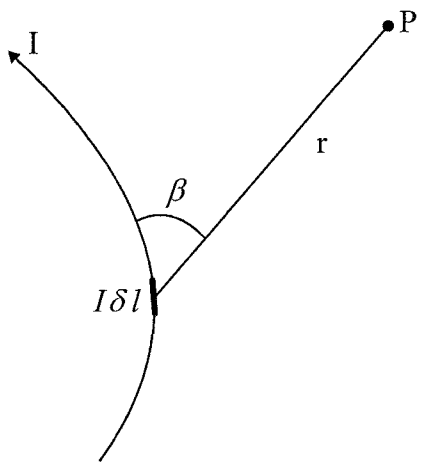


Fig 4.1: Field at P due to a single current element

To calculate the total field produced by the generator coil, equation 4.1 must be integrated around the coil perimeter. A square coil rather than a round one was chosen as the magnetic field source, since it simplifies the mathematics involved. Also, orthogonal coil assemblies are easier to construct using square coils rather than round coils (see chapter 5).

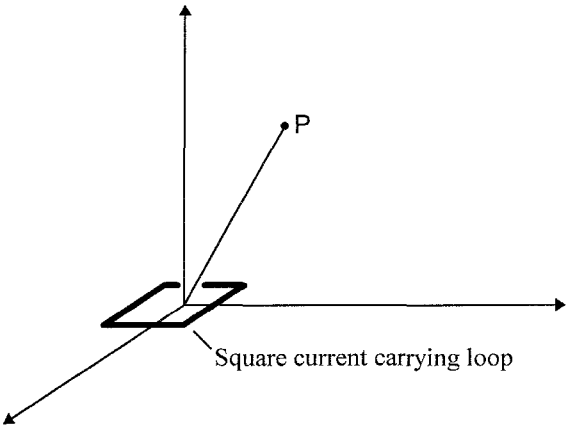


Fig 4.2: Field at P due to a square current loop. The field may be calculated by summing the field contributions from each of the four sides of the coil.

Consider a single side of the loop:

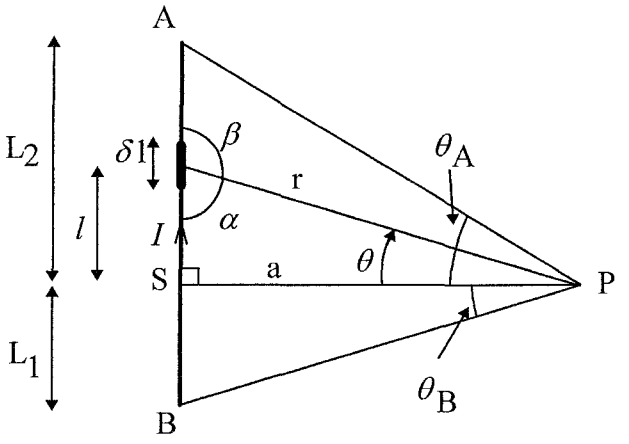


Fig 4.3: Field at P due to current carrying conductor AB
(β, r, I and δl are as in Biot-Savarts law)

To calculate the field at P we need to integrate along AB, and to do that, r and β must be written in terms of l :

$$\sin\beta = \sin\alpha = \frac{a}{r} \tag{4.2}$$

$$r^2 = a^2 + l^2 \tag{4.3}$$

Thus:

$$\delta B = \frac{\mu_0 a I}{4\pi\left(a^2 + l^2\right)^{3/2}} \delta l \tag{4.4}$$

Integrating this along AB, we get:

$$B = \int_{-L1}^{L2} \frac{\mu_0 a I}{4\pi\left(a^2 + l^2\right)^{3/2}} dl \tag{4.5}$$

Making the substituting $l = a\tan\theta$ enables this to be solved:

$$l = a\tan\theta \tag{4.6}$$

$$dl = a\sec^2\theta\, d\theta \tag{4.7}$$

The limits must be replaced by their equivalent in terms of theta:

$$-L_1 \rightarrow -\theta_1 \tag{4.8}$$

$$L_2 \rightarrow \theta_2 \tag{4.9}$$

Also from figure 4.3 and equation 4.3:

$$\frac{a^3}{\left(a^2 + l^2\right)^{3/2}} = \frac{a^3}{r^3} = \cos^3\theta \tag{4.10}$$

Substituting these into equation 4.5 gives:

$$B = \frac{\mu_0 I}{4\pi a} \int_{-\theta_1}^{\theta_2} \cos\theta\, d\theta \tag{4.11}$$

Therefore the field at B due the wire is:

$$B = \frac{\mu_0 I}{4\pi a}\left(\sin\theta_1 + \sin\theta_2\right) \tag{4.12}$$

Therefore the field at B due the wire is:

$$B = \frac{\mu_0 I}{4\pi a}(\sin\theta_1 + \sin\theta_2) \tag{4.12}$$

The direction of the B field is given by the right hand rule and so for the example above the field direction is into the page. Since B represents the field produced by only one side of the square coil, it is necessary to calculate the three values of B for the other sides and sum the results. All trigonometry must be done in 3D and the four values of B must be added using vector addition. Although it is difficult to write a single mathematical expression for the result, it is straightforward to write a computer program to calculate it. If the loop has more than one turn, then the field must be multiplied by N, where N is the number of turns.

This result applies to the case where the coil is air cored. If the coil has a permeability of $\mu_0\mu_r$ rather than μ_0, then the field will be quite different and cannot be calculated by merely replacing μ_0 by $\mu_0\mu_r$ as is sometimes assumed (this would solve the problem where the entire space had a permeability of $\mu_0\mu_r$, not just the core of the coil). Instead, numerical methods would probably be required to produce a solution.

4.2 Emf Induced In The Sensor

The emf induced in the sensor is calculated using the dot product between the field **B** at the sensor and the unit vector $\hat{\mathbf{a}}_S$ along the sensor axis, as shown in equation 4.13. k_S is the calibration constant of the sensor (Volts per Tesla).

$$V_S = k_S(\mathbf{B}\cdot\hat{\mathbf{a}}_S) \tag{4.13}$$

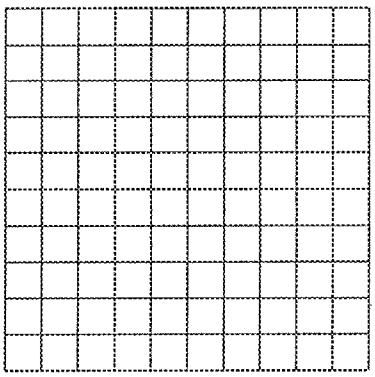
This calculation assumes that the sensor is small compared with the distance between the generator coils and the sensor, as is the case in this application. A large sensor will integrate the field over a volume rather than measuring the field at a point. The other assumption is that

the presence of the sensor in the field does not alter the field. There are two ways this could happen. Firstly, the sensor could contain a core of high permeability material to improve sensitivity. This will distort the field near to the sensor. Secondly, if the sensor coil is driving a load (for example, in the sensor amplifier), there will be a current flowing in the coil. This current will produce its own field in the sensor which will oppose the field produced by the generators and modify the field near to the sensor. Both these effects are likely to be small and are ignored here.

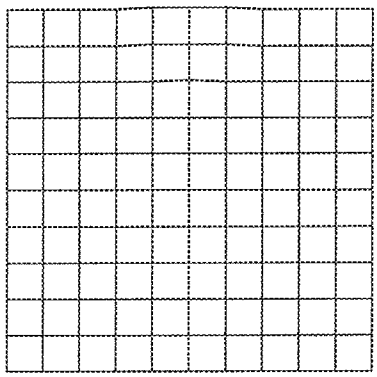
4.3 Using The Field Calculated For Practical Generator Coils To Predict The Accuracy Of The Positioning System

To calculate the accuracy of the algorithms with practical generator coils, a program called 'MAGCALC', which runs under Microsoft Windows, was specially written. MAGCALC enables a wide range of tests to be carried out on the algorithm and is described in greater detail in chapter 10. The position location algorithms, the field equations (and also the hardware characteristics) are integrated into the program, so that the entire imaging system can be simulated. The results from this program can then be compared with the results obtained from the actual system. Of importance here is the fact that MAGCALC can plot the positional output from either algorithm for different designs of generator coils. By changing the generator design and replotting the results it is possible to optimise the design for any given application.

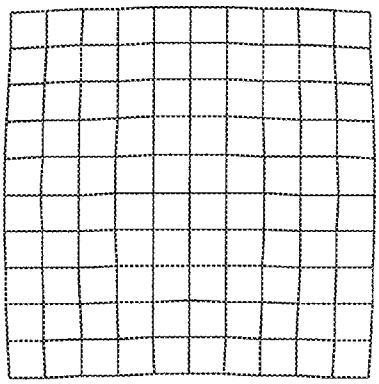
Figure 4.4 shows the calculated positional accuracy of algorithm A for four different designs of generator coil. The generator coils are all square, but have different sizes. The grid of results represents a 2D scan of 0.5 m by 0.5 m, carried out 0.4 m above the reference plane. The sensor is assumed to be pointing in the +z direction and the calculations assume that there are no errors in measuring the signals from the sensor.



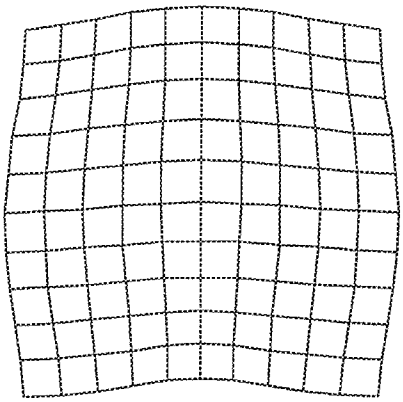
(a) Size = 20 mm



(b) Size = 50 mm



(c) Size = 100 mm



(d) Size = 200 mm

Fig 4.4: Positional results for different sizes of generator coil using algorithm A

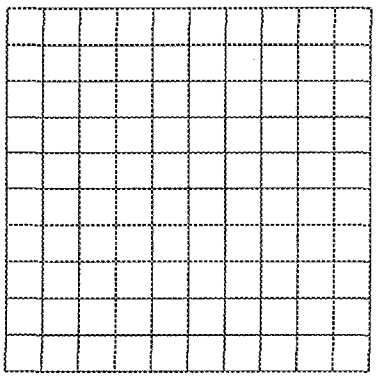
Each diagram is equivalent to 11×11 or 121 position measurements of the sensor. By joining the points to form a grid it is easy to visualise the distortion that arises as the generator size is increased. Up to 100 mm (c) the accuracy is good, whereas by 200 mm (d) the distortion is becoming significant.

The errors can be summarised as follows:

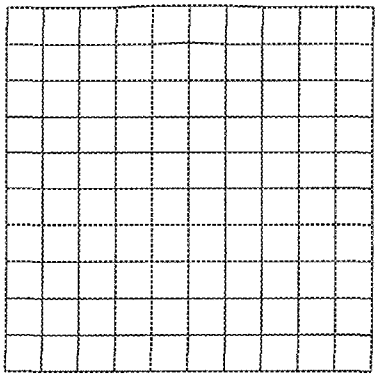
Coil Size (mm)	Max. Error (mm)	RMS Error (mm)
20	0.27	0.15
50	1.7	0.92
100	6.7	3.7
200	27	15

Fig 4.5: Table of positional errors for algorithm A.
z=0.4 m.

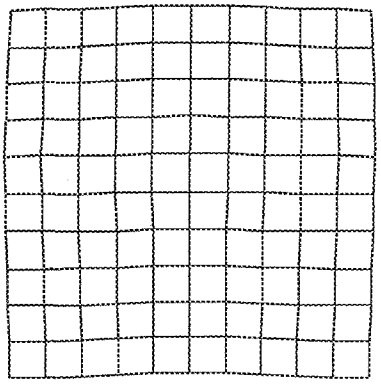
Figure 4.6 shows an identical test to figure 4.4, except that it uses algorithm B:



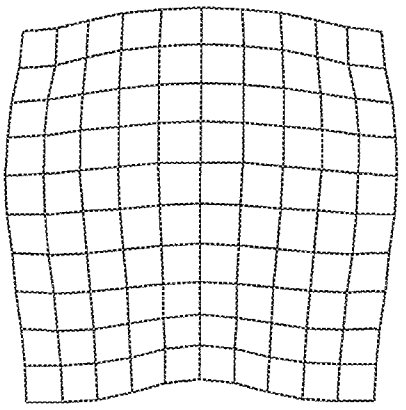
(a) Size = 20 mm



(b) Size = 50 mm



(c) Size = 100 mm



(d) Size = 200 mm

Fig 4.6: Positional results for different sizes of generator coil using algorithm B

The errors can be summarised as follows:

Coil Size (mm)	Max. Error (mm)	RMS Error (mm)
20	0.024	0.17
50	1.5	0.97
100	5.9	3.9
200	24	16

Fig 4.7: Table of positional errors for algorithm B.
z=0.4 m.

There are very few differences between these two sets of results. The maximum error of algorithm B is marginally smaller than algorithm A, but more importantly, algorithm B is considerably faster.

4.4 The Effects Of Measurement Errors

There are two types of measurement errors that affect the algorithm: random noise and signal offsets. The small size of the sensor leads to poor field sensitivity and this can result in measurement noise becoming a problem. The effects of random noise can be overcome by reducing the bandwidth of the filter which processes the signals from the sensor, though this increases the data acquisition time. Signal offsets are more serious since they cannot be removed by filtering. Offsets occur when crosstalk exists between some part of the field generation circuity and some part of the field sensing circuitry, other than by the intended route of generator coil to sensor coil. Offsets are a problem for this system, since the hardware which produces the high amplitude currents to drive the field generators is physically near the hardware which processes the small signals from the field sensors. Also, as we shall see, the algorithm is not tolerant to signal offsets.

4.5 Effect Of Random Noise On The Positional Accuracy

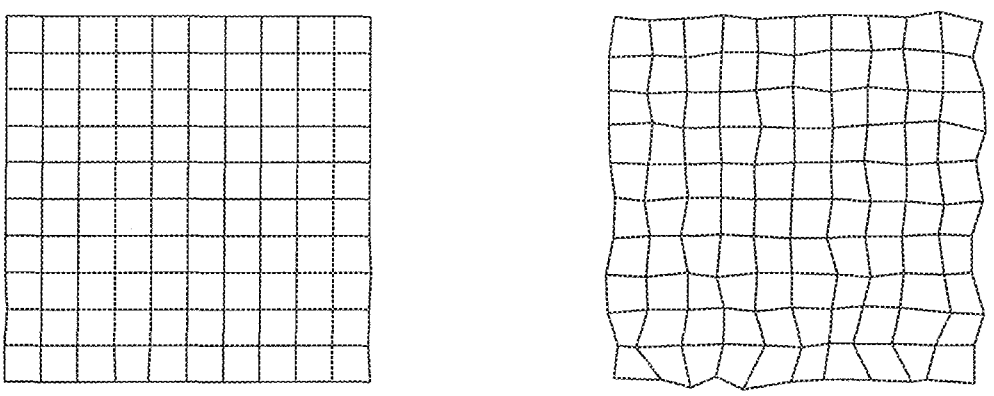
Random noise may be introduced by both the sensor device and the amplifier following the sensor. MAGCALC was used to predict the effects of this noise for a system having the following design characteristics:

Generator size:	20 mm
Generator current:	5 A peak
Generator turns:	50
Sensor sensitivity:	1×10^4 Volts/Tesla
Sensor orientation:	$\theta = 0,\ \phi = 0$

These characteristics are based on a realistic coil design, with physical and thermal limits.

The effect of adding random noise to the signal from the sensor can be seen in figure 4.8. The noise used in these simulations is not Gaussian noise, but consists of equally weighted random numbers of the range ±(Noise Amplitude) added to each sensor signal measurement.

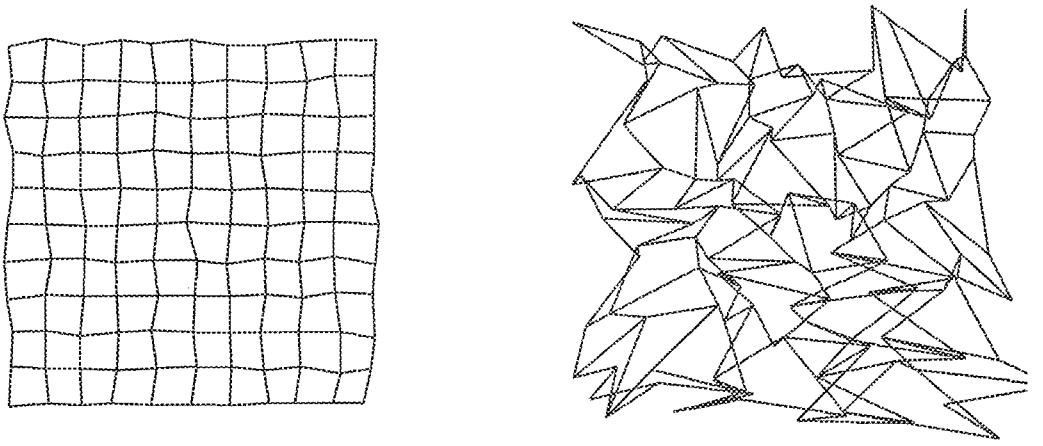
The scan shown is for x=±0.25 m, y=±0.25 m, z=0.4 m.



(a) $100\mu V$ noise (b) $1mV$ noise

Fig 4.8: Effect of signal noise on positional accuracy. z=0.4 m.

As the height of the sensor relative the generator coils is increased, the noise becomes an increasing proportion of the sensor signal. Figure 4.9 shows the same scan, but at z=0.8 m.



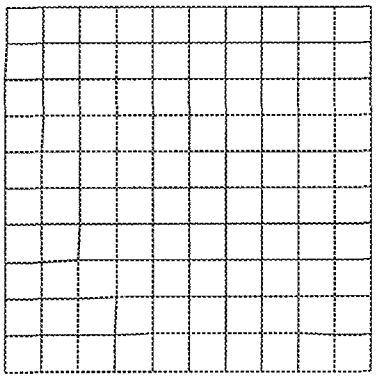
(a) $100\mu V$ noise (b) $1mV$ noise

Fig 4.9: Effect of signal noise on positional accuracy. z=0.8 m.

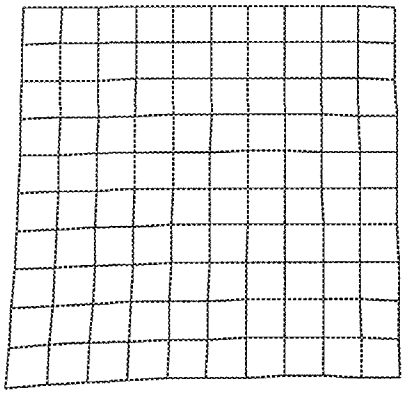
4.6 Effect Of Signal Offsets On The Positional Accuracy

Offsets occur when crosstalk exists between some part of the field generation circuity and some part of the field sensing circuitry, other than by the intended route of generator coil to sensor coil. Typically, the magnetic fields produced by the generator coils induce voltages in the sensor preamplifiers and offsets of a few μV are enough to destroy the positional accuracy of the system.

The effect of signal offsets is seen in figure 4.10 and figure 4.11. For the purpose of these calculations, the same offset is added to each of the nine field measurements used for each position calculation. This may or may not be the case in practice, depending on the cause of the offset. The system characteristics are the same as for in section 4.5.

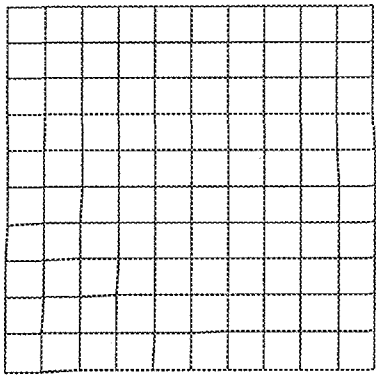


(a) $1\mu V$ offset

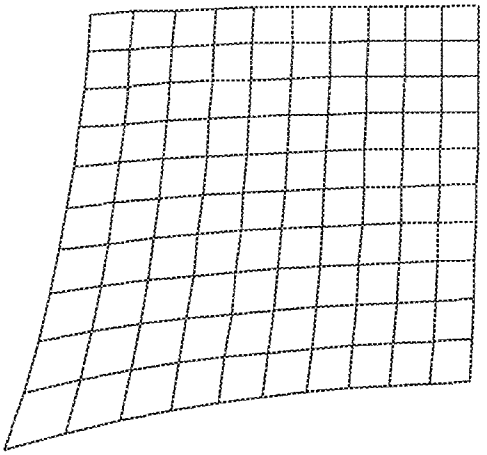


(a) $10\mu V$ offset

Fig 4.10: Effects of signal offsets. z=0.4 m.



(a) $1\mu V$ offset



(a) $10\mu V$ offset

Fig 4.11: Effects of signal offsets. z=0.8 m.

These diagrams show that even small offsets can have a marked effect on the positional accuracy. Whereas $100\mu V$ random noise produced acceptable results for a scan at z=0.8 m, a mere $10\mu V$ offset produces unacceptable results for the same scan.

4.7 Conclusion

The two position location algorithms developed in chapter 3 were developed assuming that the field generators were ideal magnetic dipoles. The most important conclusion from the tests in this chapter, is that the algorithms perform well with the fields produced by practical generator coils.

For a given signal source, increasing the size of the generator causes the field strength to increase and this is desirable since it reduces the unwanted effects of measurement noise and offsets. However, the size of the generators sets a fundamental limit to the positional accuracy of the system and this limits the maximum size of the field generators. For a target accuracy of 5 mm, figure 4.7 shows that a generator size of slightly less than 100 mm is optimum.

The position location algorithm is clearly very sensitive to signal offsets in the field measurements. Figures 4.10 and 4.11 show the effect that even small offsets have on the positional accuracy. This indicates that careful consideration should be made to minimise these offsets when designing the hardware implementation of the algorithm.

Algorithm B performs significantly faster than algorithm A, and so the later tests are done exclusively with algorithm B.

4.8 References

[4.1] B. Bolton : "Electromagnetism", Van Nostrand Reinhold (UK) (1980).

Chapter 5

Design Of The Generator Assemblies And Sensor Catheter

5.1 Generator Coil Design

The position location algorithms require three field generator assemblies, each of which comprises three orthogonal coils. The three generator assemblies must be positioned so that all the sensor coils can receive adequate signal from the generator coils over the entire positional range of the system. Also, since the generators define a plane, they must not be placed in a straight line. The accuracy of the algorithm deteriorates for small values of z (where z is measured from the plane defined by the generators to the sensor coil being measured). Therefore the generators must be placed above, below or beside the patient, and must not be mounted around the patient. Apart from these constraints, the exact spacing of the generators is relatively unimportant, provided the position location algorithm knows the position of the generators. In the prototype system the generators were attached to a wooden board and mounted just beneath the patient, so they did not obscure access to the patient by the medical staff. The accuracy of the system for different generator assembly positions can be calculated using MAGCALC.

The objective of designing the generator coils is to maximise the field produced by the coils, consistent with maintaining a field distribution which is a good approximation to the ideal dipole distribution. The latter requirement is determined by the size of the coil and was discussed in chapter 4.

To maximise the field produced by the generator coils, we need a measure of how well a particular design of coil performs. We can define a parameter F, which represents the coil efficiency as follows:

$$F = \frac{H}{\sqrt{VI}} \tag{5.1}$$

where,

V is the coil voltage (V RMS)

I is the coil current (A RMS)

H is the magnetic field strength in the centre of the coil = $k_1 NI$ (Amp Turns)

N is the number of turns

k_1 is a constant determined by the coil size

F represents the field produced by the coil for a given input power and can be used to explore the parameters which affect the performance of various coil designs.

Coil Efficiency vs Number Of Turns

$$H = k_1 NI \tag{5.2}$$

$$I = \left| \frac{V}{j\omega L + R} \right| = \frac{V}{\sqrt{\omega^2 L^2 + R^2}} \tag{5.3}$$

$$F = \frac{H}{\sqrt{VI}} \tag{5.4}$$

$$L = k_2 N^2 \tag{5.5}$$

$$R = k_3 N^2 \tag{5.6}$$

where,

H is the field strength at the centre of the coil

k_1 is a constant, determined by coil size

I is coil current (A RMS)

V is the coil voltage (V RMS)

ω is the field frequency

L is the coil inductance (H)

k_2 is a constant, determined by coil size

N is the number of turns

R is the coil resistance (Ω)

k_3 is a constant, the resistance of one turn of wire occupying all the winding volume

Substituting H, I, L and R into the expression for F:

$$F = \frac{k_1 NI}{\sqrt{VI}} \tag{5.7}$$

$$F = \frac{k_1 N\sqrt{I}}{\sqrt{V}} \tag{5.8}$$

$$F = \frac{k_1 N\sqrt{\dfrac{V}{\sqrt{\omega^2 L^2 + R^2}}}}{\sqrt{V}} \tag{5.9}$$

$$F = \frac{k_1 N}{\sqrt[4]{\omega^2 L^2 + R^2}} \tag{5.10}$$

$$F = \frac{k_1}{\sqrt[4]{\omega^2 k_2^{\,2} + k_3^{\,2}}} \tag{5.11}$$

Thus, the coil efficiency is independent of the number of turns.

Field Strength vs Drive Power

Equation 5.1 shows that to double the field strength the drive power must be quadrupled, or in other words, the power required rises as the square of the field strength. Increasing the drive power in an attempt to increase the field strength rapidly causes the thermal limits of the drive amplifiers and the generator coils to be reached.

Field Strength vs Coil Heating

The maximum field strength that a given coil design can produce is set by the maximum coil temperature. If T represents the power dissipated in the resistance of the coil, then:

$$T = I^2 R \tag{5.12}$$

$$R = k_3 N^2 \tag{5.13}$$

$$T = k_3 (NI)^2 \tag{5.14}$$

$$T = k_3 H^2 \tag{5.15}$$

where,

T is the power dissipated in the coil

I is coil current (A RMS)

R is the coil resistance (Ω)

k_3 is a constant, the resistance of one turn of wire occupying all the winding volume

N is the number of turns

H is the field strength at the centre of the coil

Thus the power dissipated in the coil is dependent of the square of the output field strength, and on the resistivity of the wire used to wind the coil. It is not dependent on the number of

turns. Another way of interpreting the result, is that for a given wire material and maximum coil temperature, the maximum field is set by the volume of the winding. Changing the wire thickness merely changes the impedance of the coil and hence the drive requirements.

For an air cored coil, there is an additional power loss due to eddy currents in the winding wire. This effect is most noticeable if the thickness of the winding is significant compared with the radius of the coil. The effect occurs because, without a core, the flux is not constrained to the centre of the coil and tends to branch out through the winding, causing induced currents in the copper. This effect increases with wire thickness and leads to the paradoxical result that in some cases, increasing the wire thickness to reduce the coil resistance, actually increases the power dissipation in the coil. The effect is difficult to calculate and is best found by experiment.

In the position location system, the nine generator coils are driven in sequence and so each coil is only switched on for a ninth of the time. This reduces the power dissipation in each coil by a factor of nine.

Generator Coil Size

Increasing the size of the field generator coils increases the field strength, but also reduces the accuracy of the field as compared to the theoretical ideal dipole as discussed in chapter 4. A square generator coil of side 50 mm is predicted to give a worst case positional error of ± 1.5 mm, whereas a coil of 100 mm gives a predicted accuracy of about ± 6 mm. The prototype system used generator coils of size 38 mm since coils of this size were convenient to produce.

Coil Core Material

Using a core material of high permeability can increase the coil efficiency, depending on the type and shape of the core. However, there are also some disadvantages. Core materials, especially ferrite, saturate at a certain field strength. This causes the signal wave form to become distorted, and since the effect changes with core temperature, causes calibration problems as the coils warm up during use.

Generator Coil Drive Circuit

An inductor carrying an sinusoidal current does not dissipate power but merely stores energy in the form of a magnetic field. Since the fields used by this system are not radiation fields, the net power into a generator coil is zero (ignoring heat loss in the winding resistance). However, this does not mean that zero power is required to drive the circuit. The voltage and current in an inductor are 90 degrees out of phase, so maximum current occurs when it has zero volts across its terminals. If the coil is connected directly to an amplifier, this causes all the power which enters the inductor to end up in the heat sink of the amplifier. The efficiency can be improved by using a capacitor to store the energy from the inductor, before returning it on the next half cycle. This arrangement is more efficient, since less power is wasted. Some power is still required to replace the losses in the system due to the resistance of the coil winding.

The simplest way to drive the generator coils is to connect each coil directly to an amplifier producing a sine wave signal. Since the only reactive component in the circuit is the inductance of the coil, this arrangement has a 1st order response.

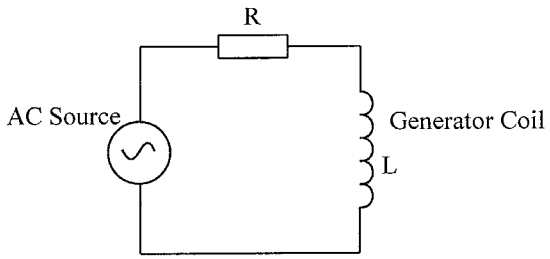


Fig 5.4: 1st order drive circuit

The drive power requirements for a given coil can be determined using the previous equations. Although it has been established that the number of turns does not affect the coil efficiency, it is still important to match the coil to a given drive amplifier to make best use of the amplifiers capabilities. For example, the prototype system used MOSFET amplifiers which can provide 30 volts peak at 5 amps peak. To match this, the coils were wound using 50 turns of 0.45 mm wire on a 38 mm diameter square core. This resulted in an inductance of about $80\mu H$ and therefore an impedance at 10 kHz of 6.3 ohms. At 30 volts this corresponds to a peak current of 4.8 amps.

The problem with a first order drive circuit is that the inductance of the coil limits the maximum current that will flow in the coil. Ignoring the winding resistance, the field for a given number of turns is given by:

$$I = \frac{V}{j\omega L} \tag{5.16}$$

$$H = k_1 NI \tag{5.17}$$

$$L = k_2 N^2 \tag{5.18}$$

So,

$$H = \frac{k_1 NV}{\omega L} \tag{5.19}$$

$$H = \frac{k_1 V}{\omega k_2 N} \tag{5.20}$$

Therefore, for a given drive voltage, reducing the number of turns increases the field. It also increases the current, and the current is limited by the maximum permissible coil temperature and the maximum amplifier drive current.

An alternative arrangement to the first order circuit is to use a capacitor to resonate the generator coil at the drive frequency, effectively nulling out the inductance of the coil. This produces a 2nd order circuit. Either parallel or series resonance can be used. A parallel circuit requires a high voltage drive at low current, and a series resonance requires a low voltage drive at high current. Apart from this, the coil efficiencies and rise times of the circuits are identical. The series arrangement is as follows:

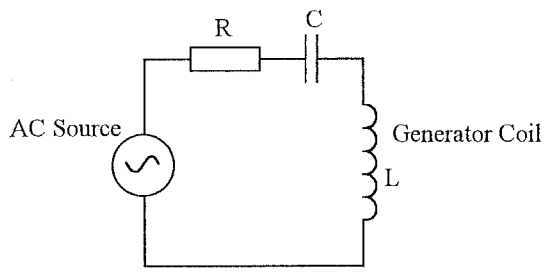


Fig 5.5: 2nd order drive circuit, using a series circuit

The coil current is given by:

$$I = \frac{V}{R + \frac{1}{j\omega C} + j\omega L} \tag{5.21}$$

The resonant frequency occurs when

$$\frac{1}{j\omega C} + j\omega L = 0 \tag{5.22}$$

or,

$$\omega = \sqrt{\frac{1}{LC}} \tag{5.23}$$

The Q of a second order circuit is given by:

$$Q = \frac{\omega L}{R} = \frac{V_L}{V_{IN}} \tag{5.24}$$

where V_L is the voltage across the inductor, and V_{IN} is the voltage across the whole circuit.

Previously, the coil efficiency F was calculated in terms of the coil voltage and current. We now call this value $F_{\text{1st order}}$. For the second order circuit the coil efficiency $F_{\text{2nd order}}$ can be defined in terms of the input voltage and the coil current.

$$F_{\text{1st order}} = \frac{H}{\sqrt{V_L I}} \tag{5.25}$$

$$F_{\text{2nd order}} = \frac{H}{\sqrt{V_{\text{IN}} I}} \tag{5.26}$$

From equations 5.24, 5.25 and 5.26, we can derive the ratio of the efficiency of a second order circuit with respect to a first order circuit:

$$\frac{F_{\text{2nd order}}}{F_{\text{1st order}}} = \sqrt{\frac{V_L}{V_{\text{IN}}}} = \sqrt{Q} \tag{5.27}$$

Thus the second order circuit has a coil efficiency of $\sqrt{Q}$ times that of the first order circuit. Put another way, only 1/Q of the input power is required to produce the same field, since the field strength is proportional to the square root of the drive power.

From this analysis, it could be concluded that Q should be made as large as possible. If this is done, however, the circuit takes a long time to settle after the drive signal is applied. Keeping the circuit in tune also becomes a problem, since slight component drift causes the coil efficiency to drop dramatically. This also means that circuitry which continuously measures the phase of the drive current is necessary, since the phase will change markedly with a small

change in the tuning of the circuit; for example as the components warm up in use. With the first order circuit the phase need be measured only once during production, since it will not change significantly after that.

The current in a second order resonant circuit order circuit, switched on at time $t = 0$, is given by:

$$I = I_o\left(1 - \frac{e^{-\xi\omega t}}{\sqrt{1-\xi^2}}\sin\left(\omega\sqrt{1-\xi^2}\right)t + \phi\right) \tag{5.28}$$

where,

I is the coil current

I_o is the current after settling

ξ is equal to $\frac{1}{2Q}$

t is time

ϕ is an angle depending on the phase of the drive signal at $t = 0$

For Q=10, the settling time to 99.8% is 20 cycles, or 2 ms at 10 kHz.

5.2 Construction Of The Generator Coil Assemblies

The generator coil assemblies used in the prototype system consist of three coils, each comprising 50 turns of 0.45 mm wire, wound on opposite faces of a 38 mm cube shaped former. The formers were machined out of Nylatron to a tolerance of 0.1 mm and 0.1 °. A major difficulty in winding the coils is ensuring that the coils are precisely orthogonal. Also, all three coils must have the same centre. In production, special cores could be manufactured, and special winding techniques could be used, but this was not possible for the prototype system. The coils were wound by hand, ensuring that each turn was parallel to the sides of the cube

and once the coils were wound the assembly was set in thermally conductive epoxy resin to protect the windings. Since two of the three windings are wound over the first winding, the three coils are not all the same size. Although this means that the coils produce different field strengths for a given drive current, the effect can be removed by software calibration.

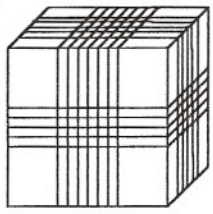


Fig 5.6 : Generator coil assembly, showing the three orthogonal coils

To reduce calibration problems due to tuning errors, and to prevent an excessive rise time for the field, the generator coils were driven using a first order drive circuit. The disadvantage of using a first order circuit is that since the power amplifier is driving a pure inductance, all the power produced by the amplifier ends up in its heat sink, producing cooling problems. However, the maximum field strength obtainable by a given coil, as limited by the heating effect in the coil, is not affected.

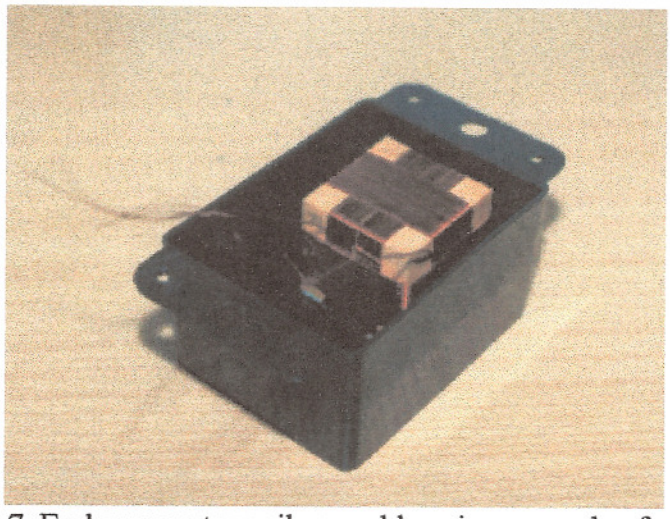


Fig 5.7: Early generator coil assembly using a wooden former.
Later versions use a plastic core

5.3 Sensor Coil Design

The biopsy channel of a colonoscope is typically 3.7 mm diameter and the maximum catheter diameter which will pass easily along this is about 3.0 mm. Allowing a minimum wall thickness for the catheter of 0.3 mm, this leaves 2.4 mm for both the sensors and the wiring. The prototype sensor consisted of a coil of wire, about 1.2 mm in diameter and 10 mm in length. The maximum length is limited by the bending of the catheter. If the sensor coil is made too long it will be damaged when the catheter is inserted into the colonoscope.

To maximise the sensitivity of the sensor, only the sensor length, number of turns and core material can be varied, since the size is fixed by the application. Although the sensitivity of a sensor coil rises with frequency, the upper limit for the frequency is fixed by other factors.

Sensor Winding

The voltage induced in such a coil when placed in a time variant magnetic field is given by Faradays law:

$$= N\frac{d\phi}{dt} \tag{5.29}$$

where,

ϕ = Magnetic flux through sensor

N = Number of turns

V = Induced voltage

The sensitivity of the coil can be increased by increasing the number of turns, since the induced voltage is proportional to the number of turns. In order to keep the volume of the winding constant this means using finer wire. Making the wire too thin causes problems, since fine wire tends to break during production when the sensors are drawn into the PTFE catheter, and could also break during service. The inductance, resistance and internal

capacitance of the coil rise with the number of turns and this can cause problems in driving the twisted pair cable and the input amplifier. Coil and cable capacitance can cause resonance problems: one prototype sensor was found to be more sensitive to Radio 4 on 198 kHz than to the 10 kHz magnetic field. High impedance lines also increases the crosstalk between different sensor channels. To prevent resonance, and to reduce inter-channel crosstalk, the wires from the coils are terminated by a resistor at the input amplifier, damping the Q of the resonant circuit caused by the sensor coil and parasitic capacitance.

An alternative way to increase the number of turns, without using finer wire, is by making the sensor longer, giving more room for the turns. The limit to how long the sensor can be is set by how much the sensor tolerates being bent, since the catheter, when inside the colonoscope, may experience bending down to a radius of curvature of 50 mm. Where ferrite cores were tried, a core of 12 mm in length was found to break in service. Other core materials, such as Mu-metal are less likely to fracture. The length is still limited to about 12-14 mm, since otherwise it can be difficult inserting the sensor catheter into the biopsy channel.

Sensor Core Material

Increasing the flux passing through a given coil will also increase its sensitivity. Placing a high permeability core into the coil has the effect of diverting flux near to the sensor to be diverted through the core, giving greater sensitivity. Although the field is modified slightly by doing this, the effect of the positional accuracy is negligible. The advantage of using a high permeability core increases with the length to width ratio. A short fat core has virtually no effect on the sensor sensitivity, whereas a long thin core can dramatically increase the sensitivity.

In the absence of magnetic sources, permeability can be described as *magnetic conductivity* since it behaves analogously to electrical conductivity. The air can be considered be have a

low magnetic field conductivity and the high permeability core a high conductivity. Inserting the core into the sensor causes the field near the sensor to change its course and go through the sensor instead of around it.

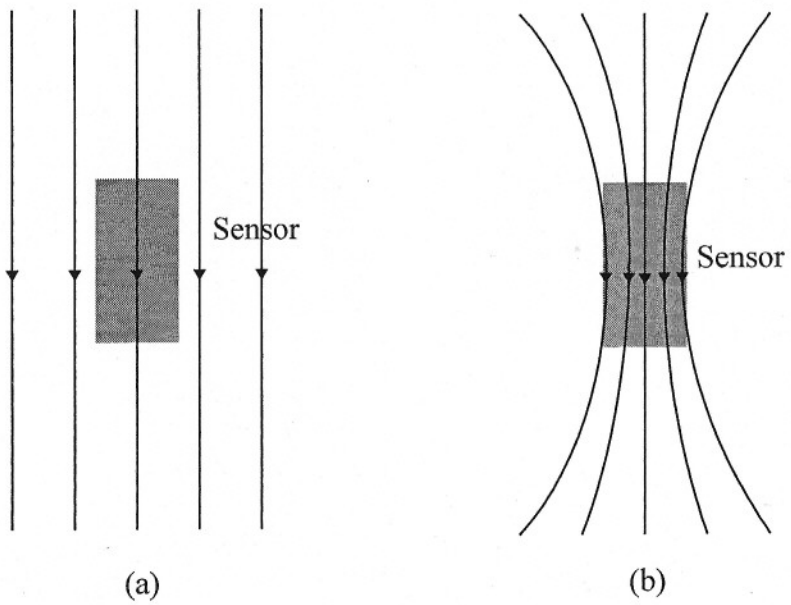


Fig 5.8: Magnetic field lines around a sensor with an air core (a) and a high permeability core (b)

The analogy between permeability and conductivity is valid because there are no sources in the vicinity. From Maxwell's equations:

$$\nabla \cdot \mathbf{B} = 0 \tag{5.30}$$

$$\nabla \times \mathbf{H} = \mathbf{J} = 0 \tag{5.31}$$

$$\mathbf{B} = \mu \mathbf{H} \tag{5.32}$$

Since $\nabla \times \mathbf{H} = 0$, we can create a scalar field, ψ, such that:

$$\mathbf{H} = \nabla \psi \tag{5.33}$$

and substituting this into the above equations gives:

$$\nabla \cdot (\mu \nabla \psi) = 0 \tag{5.34}$$

This equation is of a form which can be solved using a 3D finite difference technique. By doing this we can determine the optimum size, shape and type of the core, as well as the expected increase in sensitivity. An interactive program, called SENSOR3D was written for this purpose, and runs under Microsoft Windows. A full description of the software is included in a later chapter, but a summary of the results obtained using this program is included here. The program enables the size of the sensor and its permeability to be specified and then calculates the magnetic field in and around the sensor. The sensor is assumed to be in free air. It then uses the calculated flux density in the sensor core to calculate how much voltage would be induced in a coil wound around the core. The results have also been confirmed by experimentation.

If a sensor coil does not have a core, then the voltage induced in the coil is proportional to the area of the coil. This makes sense, since the voltage is proportional to the flux through the core, and the flux is proportional to the cross sectional area. However, the situation changes when a core is introduced. In fact, if the core permeability is high, then for a given core length, the core can be made very thin with only a small degradation in sensitivity. This is ideal for the sensors in the catheter, since the coils need to have as small a diameter as possible.

Figure 5.9 shows some results obtained using SENSOR3D, using 11 cells in the x direction, 11 cells in the y direction and 25 cells in the z direction. The core sizes are in units corresponding to the size of one cell. The size of the total volume is limited by software

restrictions of the programming language used for the program (Turbo Pascal for Windows) and the total volume to sensor volume ratio is not really large enough to enough to produce accurate results. In other words the boundaries are not really far enough away to model the realistic situation. However, the results still indicate design trends well and the small calculation volume means that the results take only a few seconds to compute.

The results (see figure 5.9) show that for a permeability of 1, the sensitivity is exactly proportional to the cross sectional area of the sensor. Increasing the area from 9 to 49, a factor of 5.4, increases the sensitivity by a corresponding factor of 5.4. For a permeability of 500, increasing the area by a factor of 5.4 only increases the sensitivity by a factor of 2.6. Thus the relationship is no longer linear. Figure 5.10 shows the results graphically. Note how the presence of a core increases the sensitivity by a factor of between 11 and 23, compared to having an air core. Increasing the permeability above 500 produces a relatively small improvement in the sensitivity.

Core Type				Sensitivity			
Width	Depth	Length	Area	μ=1	μ=10	μ=100	μ=500
3	3	20	9	7.5	53.5	146.7	173.9
3	5	20	15	12.5	83.3	200.2	229.0
5	5	20	25	20.8	127.5	268.0	297.2
7	7	20	49	40.8	219.7	395.5	452.9

Fig 5.9: Sensor sensitivity for cores of different cross section

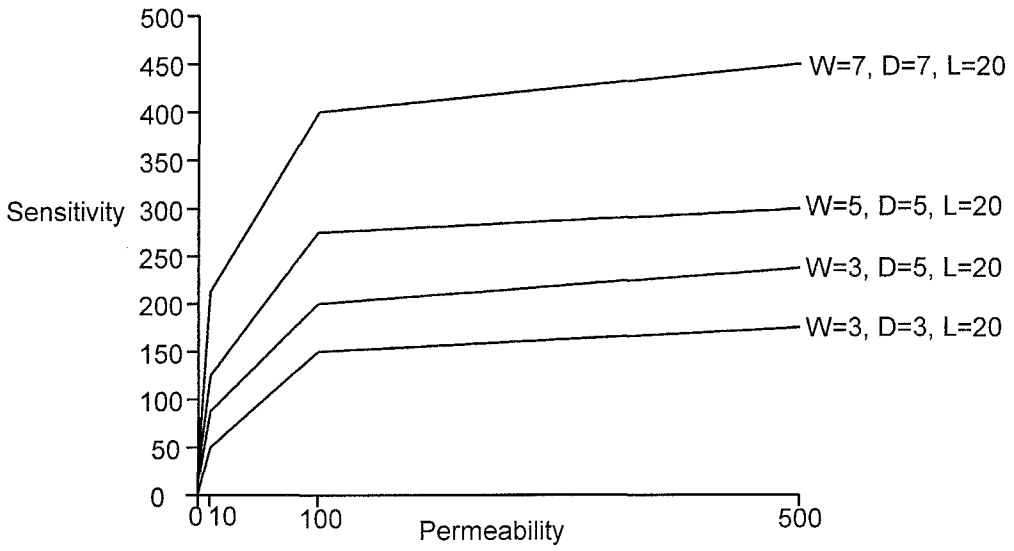


Fig 5.10: Sensor sensitivity for different cross sections

Mu-metal has a permeability between 1000 to 100000, whereas for ferrite it is only 200 to 500. The difference in signal strength is relatively small, but Mu-metal has the advantage that it does not crack or fracture when bent slightly. The key advantage in using Mu-metal is that the coil can be made flat, enabling the wire harness to pass by the sensor to the remaining sensors. Mu-metal is conductive, causing a reduction in expected sensitivity, but the eddy current losses in a flat core will most likely be less than in a cylindrical core.

The most important result from SENSOR3D is that the best way to utilise the effects of a high permeability core is to make the sensor as long as possible. Without a core, the sensitivity (assuming that the winding around the core covers the core surface) is proportional to the core length. With a core the relationship is of a higher order, see fig 5.11 and 5.13. For a permeability of 500, the sensitivity increases by a factor of 5.6 when the length is doubled from 10 to 20.

Core Type				Sensitivity			
Width	Depth	Length	Area	μ=1	μ=10	μ=100	μ=500
3	3	2	9	0.75	1.56	1.76	1.78
3	3	4	9	1.50	4.37	5.44	5.56
3	3	10	9	3.75	17.93	29.50	31.31
3	3	20	9	7.5	53.5	146.7	173.9

Fig 5.11: Sensor sensitivity for cores of different length

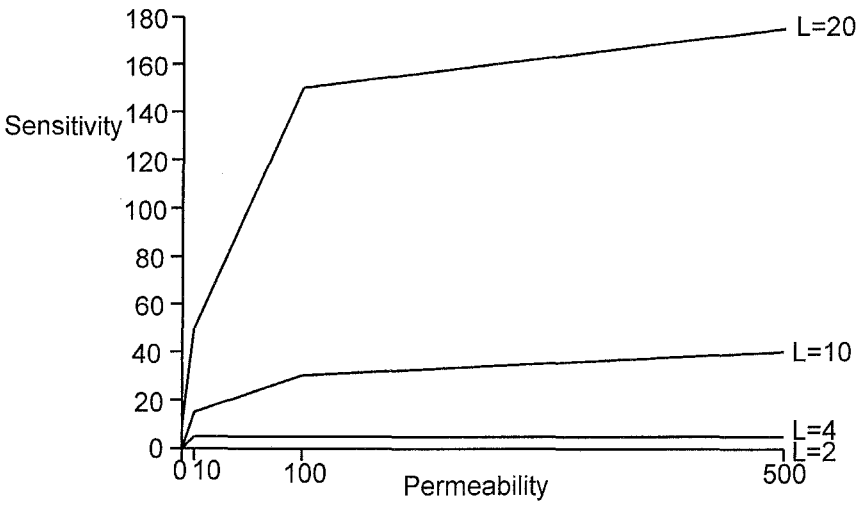


Fig 5.12: Sensor sensitivity vs core permeability for cores of different length.
Sensor width=3, depth=3

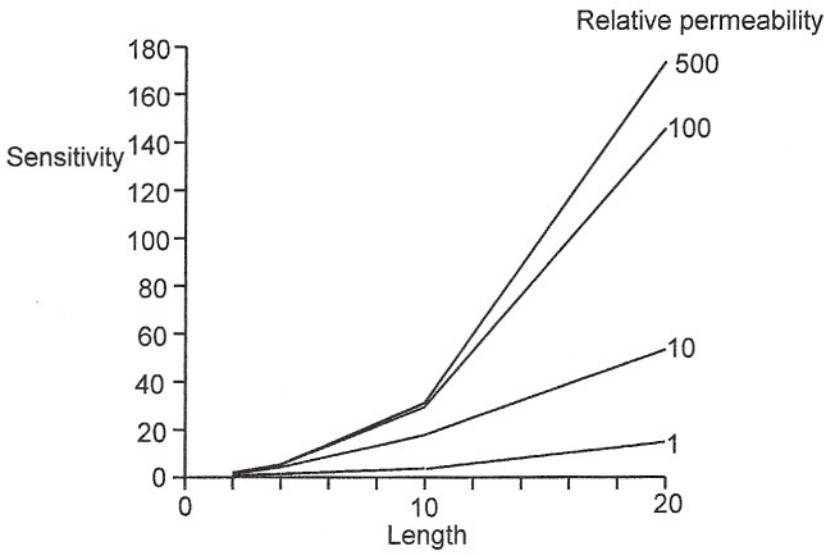


Fig 5.13: Sensor sensitivity vs sensor length for cores of different relative permeability. Sensor width=3, depth=3

5.4 Construction Of The Multi-Element Sensor Catheter

The sensor coils were wound using 6 layers of 46 swg wire on a flat Mu-metal core of dimensions 12 mm by 2 mm by 0.5 mm. To prevent the core from damaging the insulation of the wire, the core was covered with a single layer of tape. This is part of the process needs improving since it is difficult and wastes space. Once the coils were wound they were sprayed with varnish to hold the windings in place. The final dimensions of the coils are 12 mm by 2.7 mm by 1.2 mm. Figure 5.14 shows a single sensor coil.

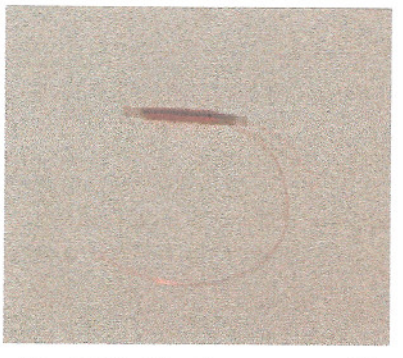


Fig 5.14: Single sensor coil

The sensor catheter consists of fourteen sensors spaced at intervals of 0.12 m along a PTFE catheter of 3 mm outside diameter. The whole catheter is 3 metres long, with sensors in the first 1.56 metres section. A medical waterproof multiway connector, mounted at one end of the catheter, connects the output from the sensors to the sampler unit. Each sensor is connected by a twisted pair of wires, making 28 wires in total. Twisted pair connections are necessary, since a common earth return wire was found to produce too much cross talk between channels. The catheter is sealed at the tip and can be sterilised by the standard colonoscopy cleaning process.

The procedure for constructing the sensor catheter is as follows:

1. Fourteen Mu-metal cores are coated with a layer of clear non-biodegradable tape.
2. 6 layers of 46 swg enamelled copper wire are wound on each sensor core to produce a set of sensor coils.
3. Fourteen pairs of 42 swg enamelled copper wires, each 3 metres long, are twisted individually.
4. The fourteen pairs of wires, and a nylon cord, are fed along the catheter leaving 1.56 metres outside.
5. The sensors are firmly attached to the wiring harness using heat shrink tape. To shrink the tape evenly, the wiring harness and the sensor are wrapped in a layer of heat shrink tape and placed inside a short piece of PTFE tubing. Hot air is then applied to the tube causing the tape to shrink evenly onto the sensor. The PTFE tube is then moved along the wiring harness to the next sensor.
6. The wires from each sensor are soldered onto a pair of wires in the wiring harness.
7. The wiring harness and the sensors are pulled into the PTFE using the nylon cord. For this operation the entire PTFE tube must be straight to reduce friction. Also the nylon cord must be kept taught from both ends.
8. The tip of the catheter is sealed with a silicone sealant.
9. The multiway connector is connected to the other end of the catheter.

This procedure is extremely tedious and time consuming. Also, it is essential that all the sensors work once the sensor is completed, so a single dry joint or broken wire renders the entire sensor useless. The PTFE tubing used to make the catheter was specially manufactured with an outside diameter of 3 mm and a wall thickness of 0.3 mm. Standard PTFE catheters used in endoscopy have a much thicker wall which does not leave enough space to allow the sensors to be drawn along the catheter during manufacture. However, the disadvantage of using a thin walled catheter is that it is prone to collapsing and damaging the sensors.

Chapter 6

A Colonoscope Imaging System Using The Position Location Algorithm

6.1 Design Objectives For The Imaging System

Early experiments showed that the original goal of range and accuracy was difficult to achieve due to poor signal to noise ratio in the signals from the sensors. The small size of the sensor coils is an essential feature of this technology but results in only very small signals being produced. A further problem is that the fields from the generators can induce signals in the wiring to the sensors and in the pre-amplifiers, and these can become a significant part of the received signal. These are referred to as signal offsets, since they cause an offset in the measured amplitude of the sensor signal. Signal offsets result in major image distortion and the effect is especially noticeable when the sensors are placed far from the generator coils, since the offsets become of comparable size to the signals from the sensors. In view of the problems of noise and offsets the system was designed to minimise these effects.

Another objective in designing the system was to make it as convenient as possible for the endoscopist. The nature of the system means that it needs to be integrated with the existing colonoscopy suite and this must be done without affecting normal procedures. The system was designed to fit seamlessly into the endoscopy environment with a minimum of inconvenience to the endoscopist, and apart from the insertion of the sensor catheter into the biopsy channel of the colonoscope and the setting of patient orientation, the system does not interfere with the endoscopy procedure in any way.

6.2 Block Diagram Of The System

The block diagram of the complete system is shown in figure 6.1. The main unit produces a high power audio frequency signal which drives the generator coil assemblies. These produce magnetic fields which induce voltages in the sensor coils mounted in a catheter inside the biopsy channel of the colonoscope. The sampler unit processes these signals and sends the phase and amplitude information to the main unit which performs the position location algorithm and generates the 3D images.

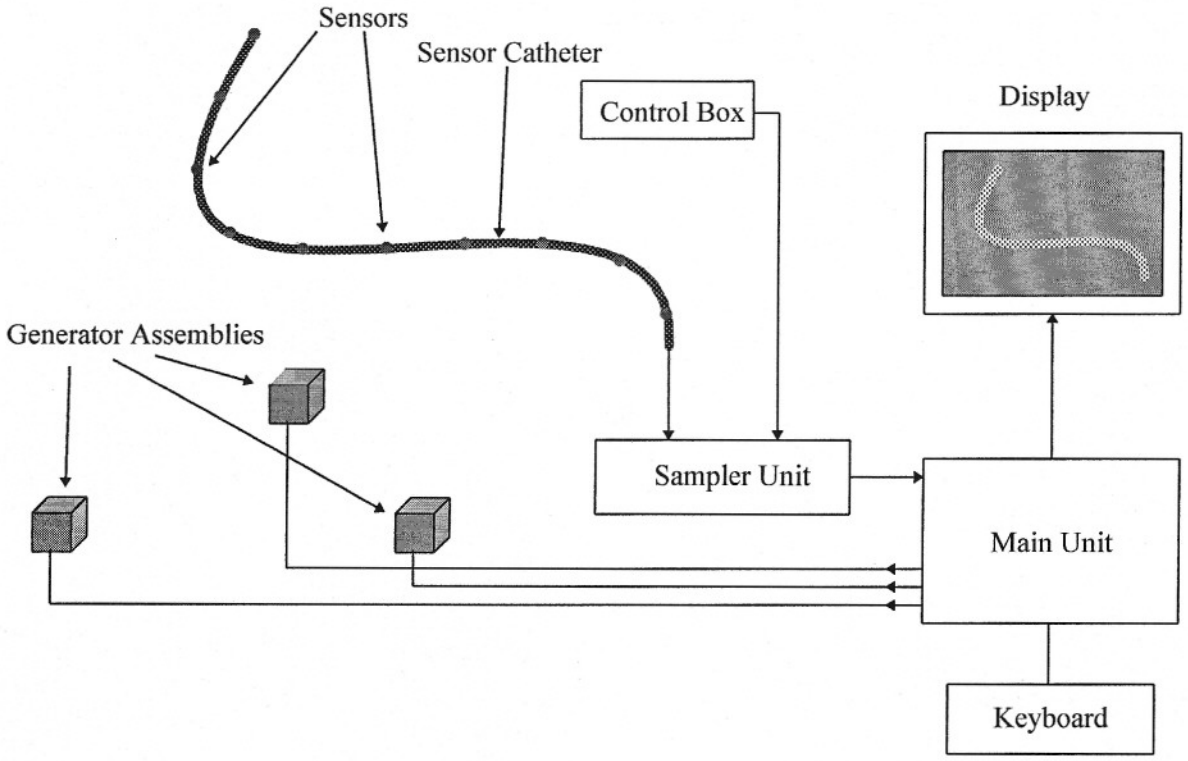


Fig 6.1: Block diagram of system

A hand held unit, called the control box, is used by the endoscopist to control the system. This box, which comprises a position sensor and miniature keyboard, is used to measure the position and orientation of the patient, and also enables most of the features of the imager to be controlled, such as rotation and scaling of the image. A full size keyboard is also available for full control of the system, though it is not necessary for normal use.

6.3 The Patient Bed

The standard bed used for colonoscopy is almost entirely metal, and was found to distort the fields unacceptably. The welded metal frame around the perimeter of the bed forms a conductive loop, which has the effect of 'blurring' the fields from the generators. The field generators induce voltages in the frame near to the generator which result in currents flowing around the whole frame. Although these currents are small, they produce a significant field because the area of the frame is large. The permeability of the metal appeared to be less of a problem and large straight pieces of steel had no significant effect on the measurement accuracy of the system.

Rather than attempting to make the system immune to the effect of the metal bed, a wooden top was constructed for use with the system. The base of the bed was retained, since it contained the mechanism for raising and lowering the patient, but the metal top was unbolted and replaced by the wooden one. The three generator coil assemblies were mounted on a wooden board which slides under the bed.

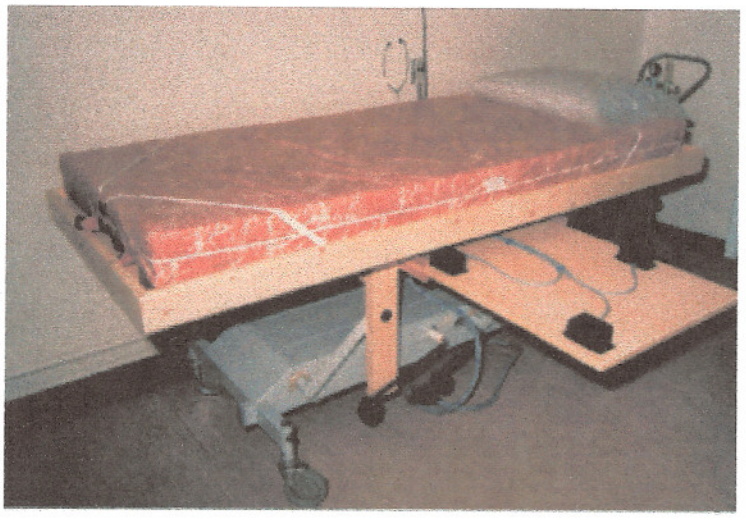


Fig 6.2: Wooden bed built for use with the imager, showing the three generator coil assemblies and the raising and lowering mechanism

6.4 Choice Of System Frequency

The system frequency is the frequency of the fields produced by the generator coils. Since the voltage induced in each sensor coil is, by Faradays law, proportional to the frequency of the magnetic field, increasing the frequency will produce a greater signal in the sensor. However, it also follows that there will be a higher voltage induced in any other conductive objects, such as the colonoscope and the metal bed. Assuming that the standard metal bed is replaced by a non-conductive bed, this leaves just the effect of the colonoscope.

The colonoscope contains a metallic braid just beneath its surface, similar to the braid inside coaxial cable. It also contains various steel wires which operate the flexible region at the tip of the endoscope. Of the two, the braiding is likely to have a greater effect on the field, since it forms a shorted turn around the sensor, whereas the wires do not form a complete electrical circuit. The braiding can be modelled by a single turn of wire which is shorted. This in turn can be represented by a series circuit containing a sinusoidal signal source, an inductor and a resistor.

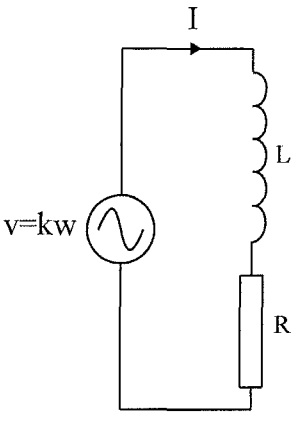


Fig 6.3: Model of the metal in the endoscope

The generator in this model represents the voltage induced in the loop and produces a signal which is proportional to the frequency of the magnetic field. L represents the inductance of the loop and R is the loop resistance. When the circuit is placed in a time variant magnetic field, a voltage is induced in the loop and a current I flows in the circuit.

The induced voltage in a coil by a time variant magnetic field is given by:

$$v = -L\frac{d\psi}{dt} = -LN\frac{d\phi}{dt} \tag{6.1}$$

Since the metal of the endoscope forms a single loop, N=1, hence:

$$v = -L\frac{d\phi}{dt} \tag{6.1}$$

For a sinusoidal field, the flux ϕ can be written as:

$$\phi = \Phi e^{j\omega t} \tag{6.2}$$

and so $\frac{d\phi}{dt}$ is given by:

$$\frac{d\phi}{dt} = j\omega\Phi e^{j\omega t} \tag{6.3}$$

The magnitude of $\frac{d\phi}{dt}$ is therefore:

$$\left|\frac{d\phi}{dt}\right| = j\omega\Phi \tag{6.4}$$

and hence V is given by:

$$= -L\left|\frac{d\phi}{dt}\right| \tag{6.5}$$

$$= -j\omega L\Phi \tag{6.6}$$

The current through the circuit is given by:

$$I = \frac{V}{j\omega L + R} \tag{6.7}$$

$$I = -\frac{j\omega L\Phi}{j\omega L + R} \tag{6.8}$$

and since $H = NI$, and N=1, H is given by:

$$H = -\frac{j\omega L\Phi}{j\omega L + R} \tag{6.9}$$

By Lenz's law, this new field H opposes the original field. It is valid to apply superposition of the fields as all the elements of the circuit are linear.

At low frequencies, the resistance of the loop dominates the impedance of the inductor ($R >> \omega L$). H is then given by:

$$H = -j\left(\frac{\omega L}{R}\right)\Phi \tag{6.10}$$

The $-j$ term indicates that the opposing field H is -90 degrees out of phase with the original field. The sum of the two fields is the field that is detected by the sensor and will have a phase shift of between 0 and -90 degrees. This is significant, since it means that the metal of the endoscope doesn't just reduce the amplitude of the field but causes a phase shift as well, and this will cause problems for the position location algorithm. The effects of phase shifts in the sensor signals are partially compensated for by inventing a method called phase compensation, which is described later.

At high frequencies, the impedance of the loop dominates the resistance ($\omega L >> R$). In this case the opposing field is -180 degrees out of phase with the original field:

$$H = -\Phi \tag{6.11}$$

Equations 6.9, 6.10 and 6.11 show that the field disturbance caused by the metal of the endoscope increases with frequency until $\omega L >> R$, when the effect levels off.

Practical experiments were carried out at 5 kHz, 10 kHz and 20 kHz to measure the positional error caused by the metal in the endoscope. At 5 kHz the distortion was minimal, but noise in the signal prevented the system producing a stable image and sufficient operational range. At 10 kHz the image distortion increased to a worst case of 25 mm, but a more stable image was obtained. At 20 kHz the sensitivity of the sensor enabled the positioning system to operate over the entire volume of interest without any difficulty. However, the worst case position error was around 50 mm.

A system frequency of 10 kHz was chosen, despite the fact that the worst case positional error at this frequency was only just within the specification for the system. Although a lower frequency would be desirable, it becomes impossible to obtain a stable image. If other means of increasing the signal can be achieved then the system accuracy could be enhanced by

reducing the system frequency as low as possible. An alternative solution to the problem would be to use a different type of sensor, instead of using a coil, which is more sensitive at the lower frequencies. Although such sensors exist, such as fluxgate magnetometers, none were found that could be miniaturised to fit inside the biopsy channel of the endoscope.

6.5 Drive Circuit For The Generator Coils

There are a total of nine generator coils which must be driven in sequence with a high current 10 kHz sine wave signal. The amplifiers chosen for the purpose were conventional 150 Watt MOSFET audio amplifiers (from Maplin Electronics) as they proved to be robust in driving a range of generator coil designs, including coils which were almost exclusively inductive. Many bipolar designs were found to fail into these loads, even with diode clamping on the output to remove the voltage spikes that arise when the coils are switched on or off. The amplifiers are mounted in the main system unit.

To enable the coils to be sequence, nine amplifiers are used, and the inputs are connected to an electronic switch which connects each amplifier in turn to the 10 kHz signal (see figure 6.4). To enable synchronous detection of the sensor signals, the 10 kHz signal must be phase and frequency locked to the sampling circuits used to measure the sensor signals, and to achieve this all the signals are derived from a single 20 MHz system clock. The 10 kHz (actually 9.766 kHz) sine wave is produced by dividing the system clock by 2048 and filtering the square wave to produce a sine wave using a 6 pole passive filter.

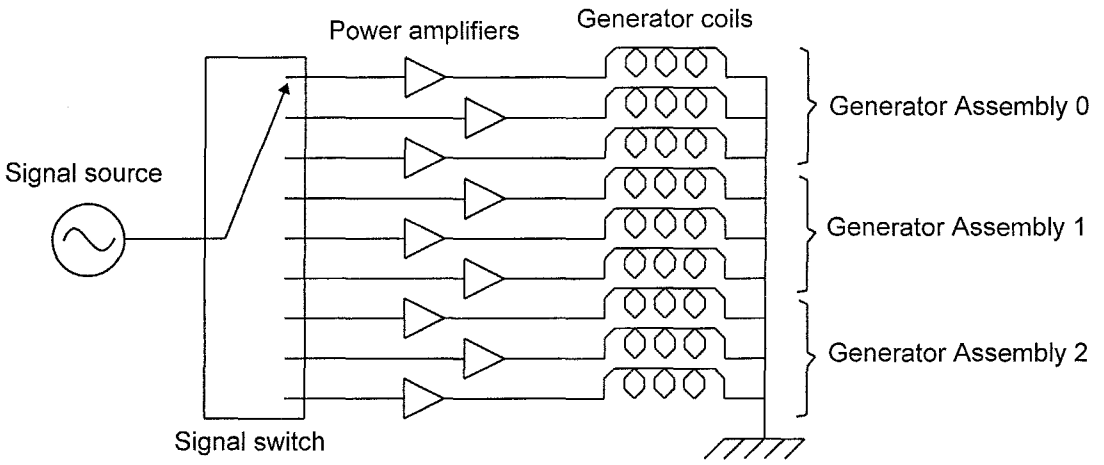


Fig 6.4: Drive circuit for the generator coils

The layout of the wiring to the power amplifiers is critical, since any loops will cause stray fields to be produced. Unfortunately, the PCBs of the power amplifiers used in the first prototype were not designed to minimise stray fields and this caused significant crosstalk problems between the power amplifiers and the sensor amplifiers. Later versions (not described here) used a carefully optimised PCB layout which eliminated the problem.

This version of the system uses a first order drive circuit for the generator coils (see chapter 5). Using a second order circuit has some advantages since it enables lower power and hence physically smaller amplifiers to be used to produce the same size fields. Although the circuitry is more complex, since the amplitude and phase of the drive current must be measured, this arrangement would enable all nine power amplifiers to be mounted on a single PCB.

Another option for driving the coils is to have a single power amplifier and to use nine triacs to switch the output. This arrangement, although extremely compact, was found to give unacceptable spikes in the drive wave form, even with negative feedback around the triacs.

6.6 Design Of The Sampler Unit

The purpose of the sampler unit is to extract the amplitude and phase components of the signals from the sensors and at the same time reject unwanted noise and signals.

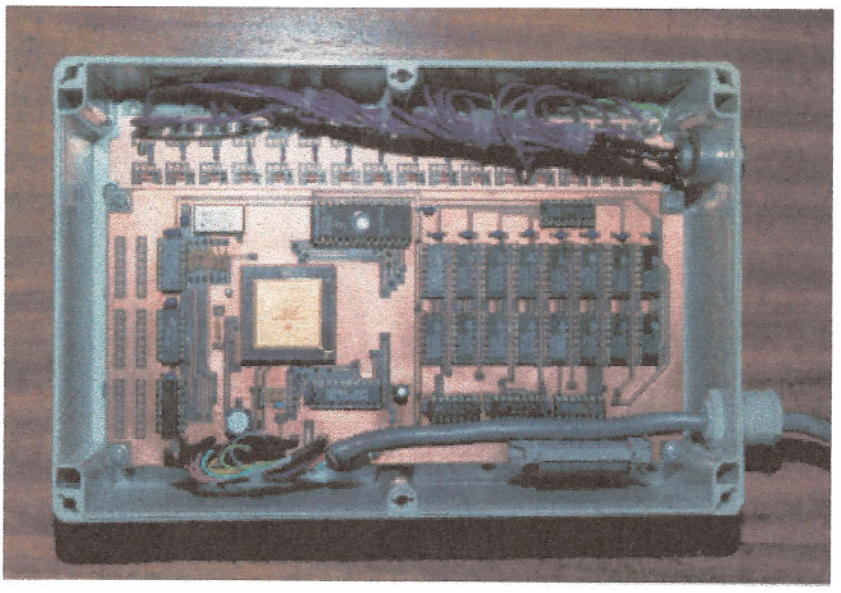


Fig 6.5: Plan view of sampler unit

The sampler has sixteen inputs, fourteen for the sensor and two miscellaneous channels. The signals from the sensor coils are small, generally of the order of a few micro-volts, and contain both random electrical noise and unwanted but coherent signals. Rejection of unwanted signals is carried out using digital bandpass filtering. Carrying out the filtering digitally has the advantage that the filter cannot drift. Conventional analogue bandpass filters produce a large phase error for a small tuning error and this causes problems for the algorithm. Also, unlike analogue filtering, the digital filtering approach does not suffer from crosstalk between channels.

The signals from the sensors are amplified by a factor of 100, and then converted into digital by 16 bit A to D converters sampling at 40 kHz. The chosen A to D converters 'over sample' at a rate of 2.56 MHz and process the samples internally to produce an effective sampling speed of 40 kHz. This technique greatly simplifies the design of the anti-aliasing filters at the

input, and as there are sixteen input channels this is a significant advantage. It is important that the sampler is not sensitive to the third harmonic of the fundamental 10 kHz frequency, since this can be produced by the generator coils, either by slight distortion of the drive signal or by using a core material in the generator coils which is non-linear. If the sampler were sensitive to this, then the positioning system would be effectively operating at 30 kHz rather than 10 kHz, making it more susceptible to eddy current effects in the colonoscope. The situation is made worse by the fact that the sensor is three times more sensitive at 30 kHz than at 10 kHz. The over sampling A to D used in this system is completely insensitive to signals at 30 kHz and so the problem does not arise. If a conventional A to D were used, sampling at 40 kHz, then the anti-aliasing filter would need a roll off of about 60 dB between 10 kHz and 30 kHz, and even then would produce an error of 0.3%. Assuming that the roll off of the filter started at 15 kHz, it would need to be at least 10th order, and would be difficult to reproduce for the 16 channels.

The sixteen digital signals are filtered in a digital signal processor (DSP) chip. The amplitude and phase are then calculated and the results sent to the main unit in the form of a serial data stream.

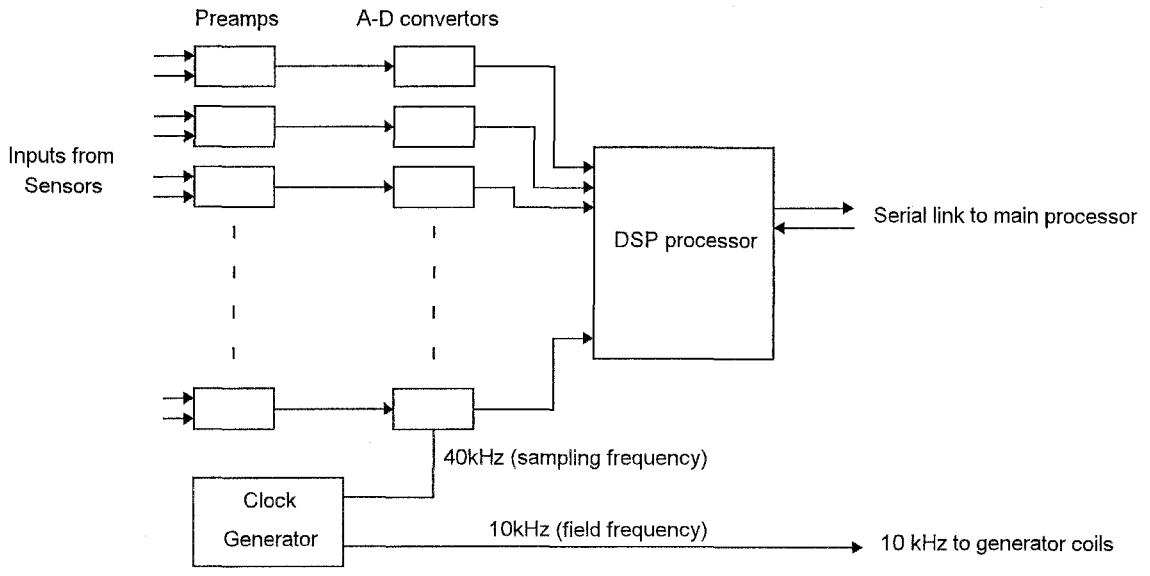


Fig 6.5: Block diagram of sampler unit.

Before the sampled data can be used, it must be windowed. The sampled data is already implicitly windowed by a rectangular window, since the sampling period is of finite length. If no further windowing were applied then the out of band rejection of the filter would only be between 20 and 40 dB, see fig 6.6.

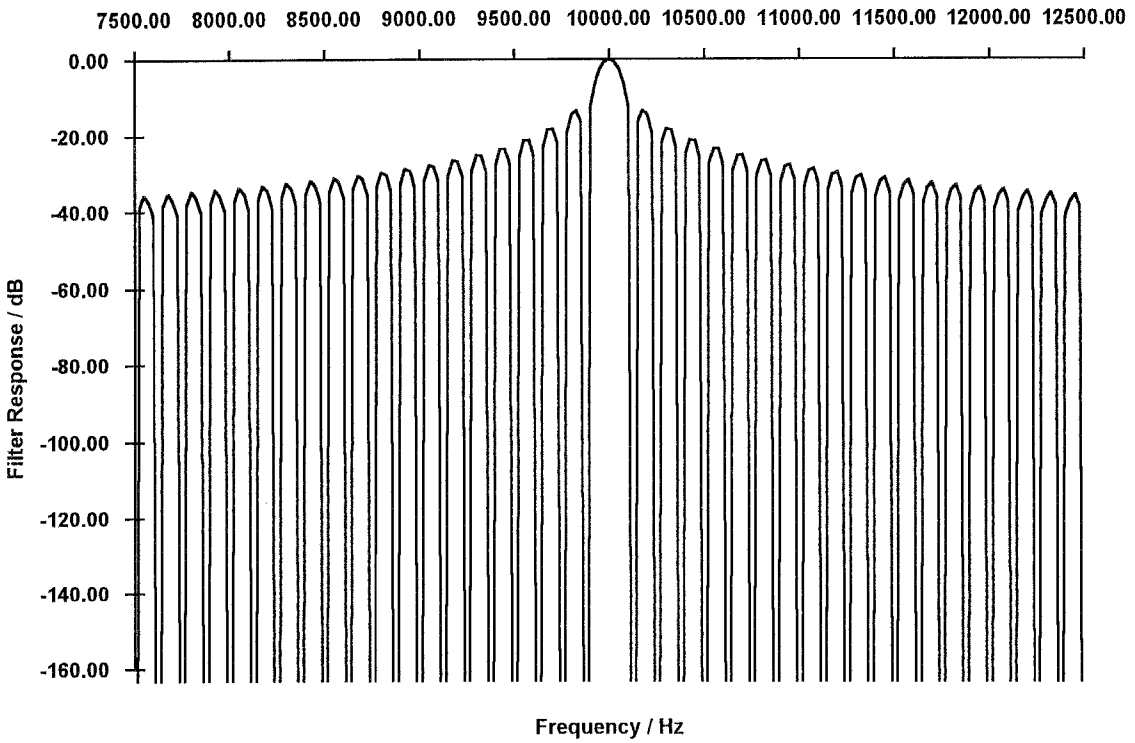


Fig 6.6: Frequency response of filter if no windowing is used

Applying a better window function, which involves multiplying each sample point by a weighting value, enables the out of band rejection to be dramatically improved. The chosen window function is the -92 dB Blackman-Harris tone decoder window [6.1]. This window is described by the function:

$$W[n] = 0.35875 - 0.48829\cos\left(\frac{2\pi n}{N}\right) + 0.14128\cos\left(\frac{4\pi n}{N}\right) - 0.01168\cos\left(\frac{6\pi n}{N}\right) \quad (6.12)$$

The Blackman-Harris window is optimised to produce a bandpass filter with low side lobes, giving excellent rejection of out of band signals. The filter is designed for use where a whole

number of cycles at the frequency of interest fits into the sampling interval. The only disadvantage of using this window in this application is that the main lobe width is high, resulting in a broad bandwidth for a given sampling time as compared with other types of window. Figure 6.7 shows the filter response for a centre frequency of 10 kHz and a sampling period of 8 ms:

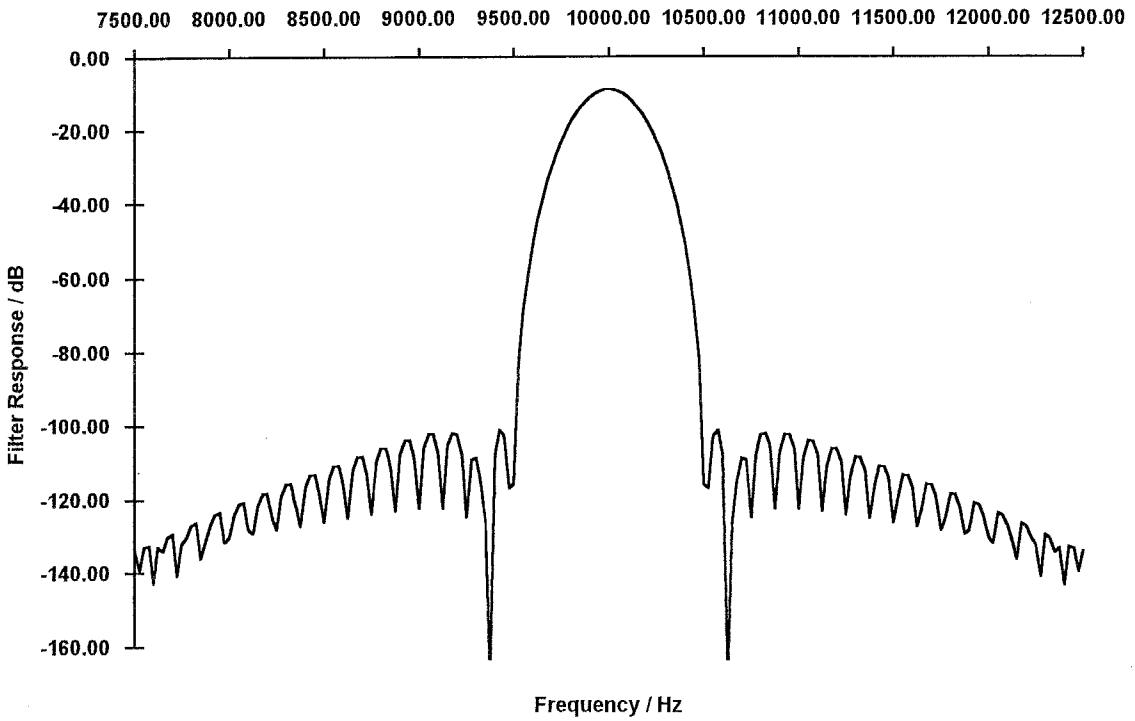


Fig 6.7: Frequency response of -92 dB Blackman-Harris window

The minimum sampling frequency is set by the Nyquist sampling criterion. This dictates that if the system frequency is 10 kHz, then the sampling frequency must be greater than 20 kHz. In this application, it is more convenient to sample at 40 kHz, since it simplifies the signal processing.

The amplitude and phase can be extracted by calculating the 10 kHz component of a discrete Fourier transform. This is identical to applying a matched filter. If $s[n]$ is the set of samples, f_r is the frequency of interest, f_s is the sampling frequency and N is the total number of samples, then F, the amplitude and phase of frequency component f_r is given by:

$$F(f_r) = \frac{2}{N}\sum_n s[n]e^{-j2\pi n\left(\frac{f_r}{f_s}\right)} \tag{6.13}$$

The result F is a complex number. Its magnitude represents the amplitude of the signal and its argument represents the phase of the signal. If f_s is set to $4f_r$, then the exponential sequence becomes 1, j, -1, -j, 1, j, -1, -j ... . The straightforward nature of this sequence means that the tone extraction becomes merely adding and subtracting the windowed samples in the correct order. The process can be summarised as follows:

- Sample N points and store in buffer.
- Window the data by multiplying each point by its corresponding value in the Blackman-Harris window function.
- Calculate F using equation 6.13.
- Calculate the magnitude of F to calculate the amplitude of the signal.
- Calculate the argument of F to calculate the phase of the signal.

This process is carried out by a DSP chip. The first three stages in the process are carried out simultaneously, since the windowing and summing is carried out cumulatively as the samples are read in from the A to D converters.

The Sampler unit consists of two PCBs and is mounted in a waterproof plastic box beneath the patient. Having the sampler unit near to the patient avoids having to feed the signals from the sensors through long cables, and is convenient for access by the endoscopist. Since it is important that the patient must not be earthed through the sensors, all the electronics in the box are isolated from mains earth using opto-isolators for the signals and a high isolation transformer for the power. This isolation circuitry is mounted in the main system unit.

One problem with designing the sampler unit is preventing stray fields from the generator coils and generator coil drive circuitry from inducing signals in the sampler unit circuitry. These

signals produce offsets in the measured sensor signal amplitudes which cause significant image distortion, as shown by the simulations in chapter 4. Careful track layout and input connector wiring reduced this to a minimum, but it is still important to mount the sampler unit both away from the main unit and from the generator coils. About 70 cm separation was found to provide adequate isolation.

The input amplifiers were built using SSM2017 differential input amplifiers, since they have ultra low noise (1.95 $nV/\sqrt{Hz}$ at a gain of 100). The gain of these amplifiers is set to 100 by a single resistor. Unfortunately these amplifiers have a gain tolerance of ±0.2 dB (±2.3%) which is not acceptable. This error is removed by a one-off software calibration.

The sixteen outputs from the preamplifiers are then filtered to removed RF signals and are then converted into a serial bitstream by eight stereo CS5338 (Crystal Semiconductors) analogue to digital converters. These over sampling converters sample the input at an effective speed of 40 kHz. The serial bitstream from these converters passes to the DSP board where the windowing and digital filtering is carried out. The DSP chip used is a DSP56001 (Motorola) running at 20 MHz.

6.7 Main System Unit

Apart from the nine power amplifiers, the main system unit contains various power supplies, the isolation circuitry for the sampler unit, and the CPU board which implements the point location algorithm. The CPU board also produces the real time 3D graphics images.

Endoscopists using the system for colonoscopy research often wish to analyse the intubation procedure after the procedure has been carried out. To enable this, a hard disk is included in the main system unit which stores the images at a rate of one per second. It was not found necessary to save every image. The images are stored in files and may either be viewed on the

imager system, or may be transferred to a PC (via a floppy disk) and viewed through a custom program called ENDOVIEW for Microsoft Windows™. Typically, between three and five image files may be saved onto a single 1.44 MB floppy disk.

6.8 Purpose And Use Of The Control Box

The control box contains a position sensor and a keypad and is used by the endoscopist for setting up the boundaries of the patient. The box is placed at three standard anatomical positions in turn - the splenic flexure, the hepatic flexure and the anus - and a button corresponding to the position is pressed. This sets three markers on the imager screen, referred to as 'static markers'. The imager software uses these points to present the endoscopist with a front view (AP or anterior-posterior) and a side view (LL or left lateral), regardless of the orientation of the patient relative to the bed. This is necessary since the patient may lie in any position during the procedure. The static markers are also of use to the endoscopist in interpreting the imager view, and determining the progress around the colon.

6.9 The Phase Compensation Method

In the absence of any unwanted conductors, the phase of the signal in the sensor will be either 0° or 180° depending on the orientation of the sensor with respect to the field. As the sensor is rotated, the amplitude will vary according to the cosine of the angle between the axis of the coil and the direction of the field at the sensor. On a phasor diagram, as the sensor is rotated, the phasor starts at $V\angle 0°$, shrinks towards zero, and then grows towards $V\angle 180°$. Since there are only two possible phase angles, we can talk about positive (phase=0°) and negative (phase=180°) voltages. In fact, the position location algorithm is not concerned with the phase of the signal and instead takes a set of signed voltages as its input.

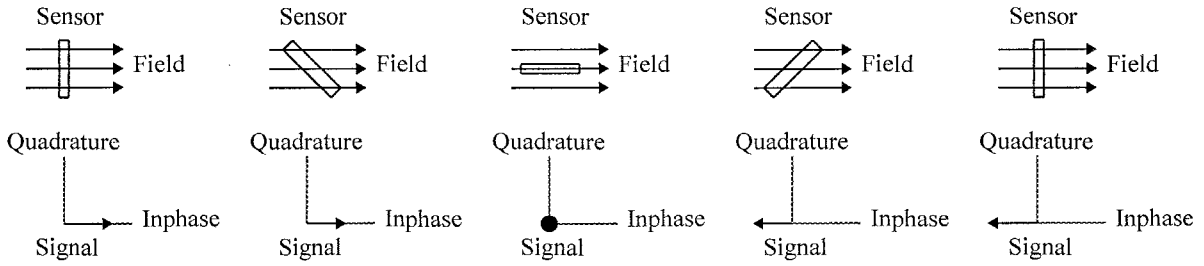


Fig 6.8: Signal in sensor for different orientations of the sensor

Note that when the sensor axis is perpendicular to the field direction, the induced voltage is zero. In the presence of a conductive loop around the sensor, however, the induced voltage (or phasor) never falls to zero, but rather traverses in an arc across from $\angle 0°$ to $\angle 180°$.

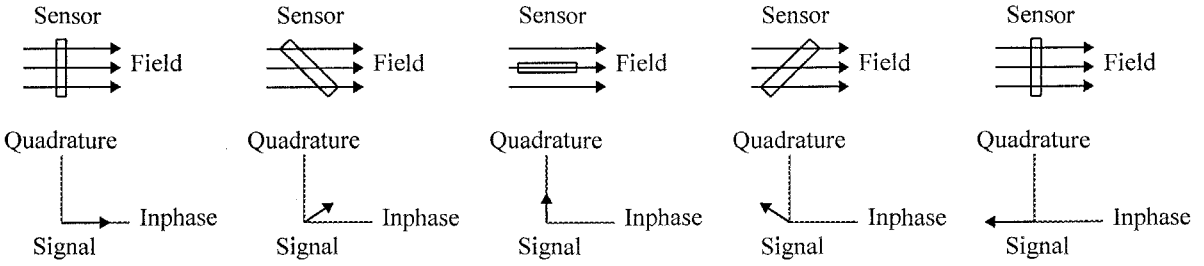


Fig 6.9: Signal in sensor with shorted loop around the sensor

The problem with this is that the sign of the sensor voltage changes from positive to negative when the amplitude of signal is not zero. The effect on the algorithm is to produce a discrete jump in the calculated position of the sensor, for a small change in the sensor orientation. Apart from being incorrect, it produces a 'lumpy' image of the colonoscope.

The solution to the visual problem, but not to the loss of accuracy, was to invent a method referred to as 'phase compensation'. This technique involves multiplying the signal from the sensor by the cosine of its phase. The method works because the unwanted signals are at a phase of $\pm 90°$. Since the cosine of $\pm 90°$ is zero, this method ensures that the phasor changes sign at zero amplitude, removing the unwanted lumps from the image. It has no effect on the image quality in the absence of any metal work. Phase compensation is a cosmetic solution to

the problem, since it improves the visual appearance of the image without necessarily making it more accurate.

$$_S' = V_S \cos\theta_S \tag{6.14}$$

If the system frequency could be operated below the point where eddy currents become significant, then phase compensation would not be necessary. If this is not possible, then phase compensation can be used to give a more presentable image.

6.10 References

[6.1] F.J. Harris : "On the use of windows for harmonic analysis with the discrete Fourier transform ", Proceedings of the IEEE, Vol. 66 No. 1, January 1978.

Chapter 7

System Software

7.1 Introduction

Software is at the heart of nearly all modern electronic systems. It allows a degree of system control and on-line decision making which was not possible before its introduction. The endoscope imaging system uses software for a variety of applications, including solving the position location algorithm, digital signal processing and 3D graphics. The software also implements a user interface to enable the endoscopist to control the system.

Medical software must be reliable. A major problem of current programming languages such as C, C++ and PASCAL, is that the sprawling text format of these languages invariably leads to a high number of software bugs. Software bugs occur either when then programmer has not solved the original problem correctly (i.e. an algorithm error) or has not programmed the software to cope with a particular eventuality. Algorithm errors are difficult to avoid, since it is difficult to prove that the original problem has been correctly analysed and solved. The problem of the program getting into a situation where it reacts incorrectly when unexpected events occur should be more solvable. Such events could be due to an external event, such as an unexpected key being pressed or a connector being disconnected, or due to an internal event, such as an unexpected result from a calculation. Existing languages, whether procedural or object oriented, do not have rigorous ways of ensuring that the program is equipped to handle all the eventualities that it will encounter. Instead, the program is written assuming that only the expected eventualities will occur, and then patched as bugs are detected. With modern programs totalling between 25000 and 200000 lines of code for standard desktop applications, bugs have become unavoidable.

The majority of the software in the main unit is written in Borland Turbo Pascal, with some graphical routines written in assembly language for speed. The DSP signal processing software in the sampler unit is written exclusively in assembly language. Pascal runs at about the same speed as C, and it may be argued that it tends to be slightly less bug prone due to stronger data type checking. In favour of C is the fact that more debugging tools are available. The main unit software currently runs under the ubiquitous Microsoft MS-DOS operating system. This is due to the history of the project, when the system was based around a stand alone IBM PC. At a later stage, the software may be made independent of a standard operating system, since those currently available for the 486 (MS-DOS / OS2 / UNIX etc.) impose significant restrictions on the machine in terms of memory organisation, display access and general complexity. They also introduce the possibility of viruses entering the machine via the floppy disk drive used for saving images. Other operating systems are available for embedded microprocessors, but these have not been explored.

The CPU board in the main unit is a conventional 486 PC motherboard. These boards are cost effective for numerically intensive applications, especially as the 486 has on chip floating point capabilities. They can also be programmed using many popular languages. A serious and potentially very time consuming problem with using a standard PC under MS-DOS is that it is based on a old 16 bit architecture. This means that the entire memory is divided into 64 kB blocks, which makes it difficult to create large items of data. This limitation is another good reason to abandon MS-DOS. The sampler unit is based on the Motorola DSP56001, which is a relatively expensive chip (£69), but requires only simple hardware.

7.2 ENDOSCAN

The software which solves the position algorithm and produces the 3D images is called ENDOSCAN. This software provides a user interface to enable easy control of the system and

manipulation of the 3D images on the screen, and also controls the sampler hardware. The software is written in a combination of Turbo Pascal and assembly language.

7.2.1 Designing The 3D Display

The most important part of the software, from the endoscopists point of view, is the graphics. The image on the screen must make best use of all available data to produce an image which is of maximum benefit to the endoscopist. In many cases, a compromise must be made between screen content and image clarity.

Early versions of the system used ray-tracing and reflectivity models to render the image of the colonoscope in 3D [7.1, 7.2]. This produced a highly attractive image, although strictly speaking not correct since a real colonoscope is not reflective. This system also used perspective to give the impression of depth. Both techniques proved to be unpopular with the doctors, since the reflectivity model was unnecessary and the perspective gave rise to 'z distortion'. z distortion means that the same image, when moved further away, appears smaller. This is not desirable for colonoscopy, since the size of loops must be instantly visible, without having to mentally compensate for the image depth.

To solve these problems a much simpler technique was implemented, where the colour of the colonoscope is determined by its depth into the screen. Thus a part of the colonoscope which is further from the viewer appears darker than a part which is nearer. Cross over regions appear clearly with this technique, see figure 7.1, and this is considered essential for unlooping knots which may form (a key advantage over x-ray techniques). The perspective was also removed, so that the image does not get smaller as the image gets further away. The result is an image which conveys all the essential information at a glance. The graphics software includes a full range of 3D image manipulation controls, including six axes of rotation and variable magnification.

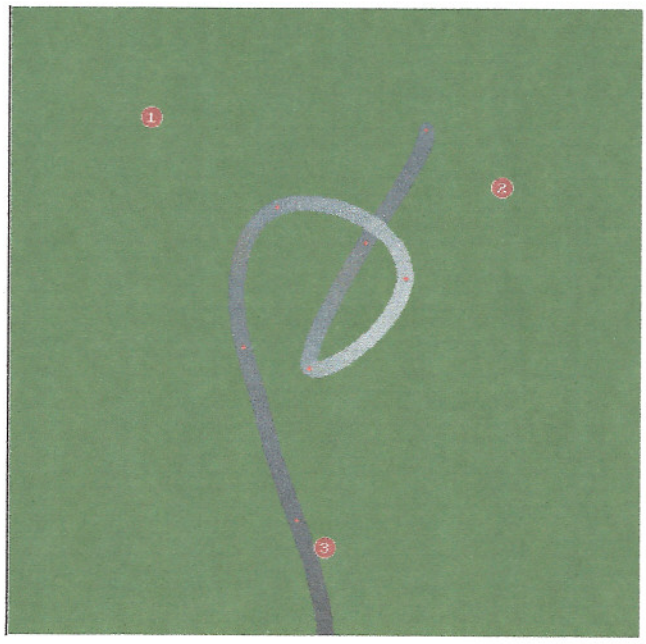


Fig 7.1: Typical imager view, showing the 3D effect

7.2.2 Developing A Curve Fitting Technique

The output from the position location algorithm is a series of points corresponding to the sensors inside the colonoscope, where each point is described by its position (x, y, z) and orientation (θ, ϕ). To create an image of the colonoscope configuration, a curve must be fitted through these points. The simplest way to do this is to fit a best fit line through the points using B splines or other curve fitting technique. This method was tried initially and it produces a visually convincing result. However, it is not too accurate since the curve is not guaranteed to go through all the measured points, and corners tend to get 'short-cut'. The solution to this problem was to use the sensor orientation information in addition to the position information and to fit a cubic equation between each sensor point. Three cubic equations are fitted, one for each of the three dimensions. The equations were solved so that the path length of the curve between consecutive points is equal to the physical path length on the colonoscope. This technique enables a curve to be fitted between any given pair of points, without requiring a knowledge of the position of any adjacent points. This is different to the B spline approach, where a knowledge of several points either side is necessary.

Consider the general parametric cubic equation, representing one axis x, y or z, and with parameter p representing the distance along the cubic (equivalent to the distance along the colonoscope).

$$f(p) = ap^3 + bp^2 + cp + d \tag{7.1}$$

$$\frac{df}{dp} = 3ap^2 + 2bp + c \tag{7.2}$$

Let the measured positions be at positions p_1 and p_2. Without loss of generality, p can be assumed to take the interval 0 to 1 ('normalised' length of colonoscope between points).

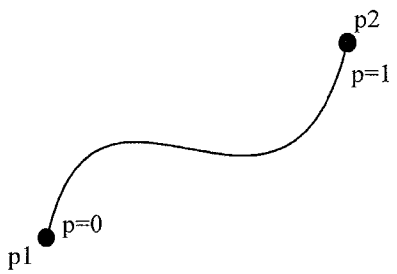


Fig 7.2: Parametric cubic between adjacent points

For the x axis,

$$f_x(0) = d \tag{7.3}$$

$$f_x(1) = a + b + c + d \tag{7.4}$$

$$\frac{df_x(0)}{dp_1} = c \tag{7.5}$$

$$\frac{df_x(1)}{dp_2} = 3a + 2b + c \tag{7.6}$$

These equations can be combined to produce the coefficients a, b, c and d:

$$a = \frac{df_x(0)}{dp} + \frac{df_x(1)}{dp} + 2\left(f_x(0) - f_x(1)\right) \tag{7.7}$$

$$b = 3\left(f_x(1) - f_x(0)\right) - 2\frac{df_x(0)}{dp} - \frac{df_x(1)}{dp} \tag{7.8}$$

$$c = \frac{df_x(0)}{dp} \tag{7.9}$$

$$d = f_x(0) \tag{7.10}$$

Substituting the coefficients into the general cubic equation, gives a cubic which describes the path in the x axis. The function values f_x are given by the measured positions and its derivatives are calculated from the measured orientation of the sensors at the two points. The equations are similarly derived for the y and z axes. By substituting values of p between 0 and 1 into the three cubic equations, values of x, y and z are obtained. In practice around ten points are obtained and straight lines are drawn between them.

This method of fitting cubic equations between measured points is only an approximation. It works reasonably well because the measured points are sufficiently close together that any curves that occur between the points cannot have an order higher than three, hence the use of a cubic equation. This is guaranteed by the mechanical construction of the colonoscope. A better solution would be to use a knowledge of the mechanical behaviour of the colonoscope to calculate the path that the colonoscope must take between two points. This could be done, for example, by simulating the behaviour of an articulated chain with springs at each junction, fitted between the two known points. The solution could be found by minimising the strain energy of the chain. Although this would be more accurate, it was felt to be unnecessarily complicated for this application, due to the computational overhead and the small advantage gained over the simpler cubic method.

7.2.3 Creating The 3D Effect

The image of the colonoscope configuration is built up from hundreds of short lines of a width which imitates the appearance of the colonoscope. Each line is created by drawing multiple circular 'blobs' whose colour corresponds to the depth of that part of the line in the viewing direction and whose diameter corresponds to the required line width. In order to create the 3D effect, the colour of each pixel of the blob is compared with the screen pixel colour, before that pixel is drawn. Only pixels of a higher number, or brighter colour, are drawn onto the screen. This means that where one part of the colonoscope passes behind another part, the part which is not visible does not get drawn.

7.2.4 Producing A Real Time Display

The graphics resolution used is 640 by 480 at 256 colours. This is a high enough resolution to prevent the images looking pixelated, without taking too long to draw. Even the most modern PC graphic cards are notoriously slow, partly because the video memory is divided into multiple blocks of 64 kB each. For each pixel drawn it is necessary to determine which block that pixel will be in, and if this is not the current block, to select it. The overheads of doing this can be considerable. An alternative, not easily done with MS-DOS present, is to use the memory management hardware of the 486 to generate page faults if an attempt is made to draw outside the current block. The page fault handler can then select the required block. This technique is used by the GNU-C/C++ compiler for the PC.

To obtain a real time display which updates the image without a noticeable delay when the colonoscope is moved, between 5 and 15 complete images or frames are required per second. The maximum frame rate for the system is currently limited to 8 frames per second by the sampling hardware. Fortunately the sampling hardware runs in parallel to the main processor, and so, allowing for data transfer between the processors, the overall rate is only slightly less than 8 frames per second.

If the screen were cleared and redrawn for each frame, unacceptable flicker would occur. To avoid this, the screen must be updated only when it is not visible to the viewer. This is called video page swapping, and requires two video pages, only one of which is visible at a time. The screen is updated by drawing to the invisible screen and then the two pages are swapped over, so that the new screen becomes visible. Since the swapping process only involves changing a flag on the video board, it takes only a negligible amount of time, and the result is a moving display with no flicker. The problem of doing this on standard PC video boards is that page swapping is not generally supported at the required video resolution of 640 by 480 by 256 colours. However, for selected boards this facility can be obtained by abandoning the standard graphics units and writing special purpose graphics commands. Although this restricts the software to a limited number of graphics cards, it is not a serious limitation in this application since the video board is housed within the system.

7.2.5 Implementing The Position Location Algorithm

The position location algorithm is a direct implementation of algorithm 'B' developed in chapter 3. It is essentially a module which takes nine measured values of field strength and converts them into five positional parameters. It also uses a knowledge of the positions of the generator coils. The field strength values that the algorithm uses are normalised, such that a dipole produces a sensor signal of 2 at 1 metre from the dipole along the dipole axis, and a sensor signal of 1 at 1 metre from the dipole perpendicular to the axis.

7.2.6 Managing System Calibration Data

After the sensors and generators have been built, they are tested to determine their sensitivity. This calibration information is stored on disk and used by the program to calibrate the signals

obtained from the sensors. A problem with this is that the program needs to be told which catheter is being used so that it can select the relevant calibration data. Future designs may store the calibration data in an EPROM or similar device, inside the connector to the sensor. Alternatively, the sensors could be manufactured to high tolerances so that standard calibration data can be used for all sensors.

7.2.7 Static Markers and Patient Orientation

The imager must display the front and side views of colonoscope, regardless of the orientation of the patient on the bed. Therefore the system needs to have a way of knowing the orientation of the patient. Many different methods of determining this were tried during the clinical trials, with the aim of minimising the inconvenience to the endoscopist. Endoscopists also requested some form of patient outline on the screen.

The first method involved positioning a sensor catheter at three standard points on the patient, and pressing a key on the keyboard for each point so that the computer would store its position. The software then uses these points to set the two viewing directions. The points also appear on the screen as small red and white circles containing the number of the marker. This method operates satisfactorily, but requires the orientation to be set again whenever the patient is moved. Certain colonoscopy techniques require the patient to be moved a number of times during the procedure.

Another approach was to use a set of standard views. Since there is a limited number of possible patient orientations, the endoscopist merely had to set the current patient position from a choice of four. When a given view was selected, a picture of the bed appeared, together with an arrow which showed the current viewing direction. This method proved unsuccessful, since it required that the patient was carefully aligned with the bed to prevent the image appearing off the screen or so dark or bright that the 3D effect did not operate

properly. There was also no screen information about the outline of the patient to help the endoscopist interpret the image.

To obtain a full outline of the patient, a sensor was mounted in a biro pen case so that it could be used to draw around the patient. Pressing a button on the pen caused the software to continuously store points as the pen was moved around the patients abdomen. This produced a reasonable image, although it was difficult to drag the pen accurately past the side of the patient nearest the bed. The main problem with this method was that it was difficult for the software to extract the viewing direction from the data. The general opinion of the endoscopists was that it was only necessary to see three strategic points, the hepatic flexure, splenic flexure and anus, rather than a complete outline which could be confusing.

The next idea used two sensors, mounted in a small plastic box, which functioned in a similar way to a normal camera. To set a given viewing direction, the box is pointed at the patient in the required direction and a button is pressed on the box. The software then uses the information from the two sensors in the box to calculate the viewing direction and display an image of the colonoscope as would be seen by an imaginary camera placed at the position of the box. For example, to obtain a front view, the box is pointed at the front of the patient and the 'set' button pressed. Although this approach is elegant, there are two disadvantages. Firstly, it does not display any information about the patients outline. Secondly, it proved impossible to explain to the endoscopists that, just like a real camera, it matters which way up the camera is held. The result was that they were getting images which were viewed from the correct direction, but which were rotated by 90 or 180 degrees. This is equivalent to holding a conventional camera upside down when taking a picture.

To attempt to keep the viewing direction correct when the patient is moved, a sensor was attached to the patient by sticky tape. When the patient moves, the software detects the movement by the change in the position and orientation of the sensor attached to the patient, and changes the viewing direction accordingly. The problem with this method is that the

wires to the sensor proved irritating since they tangled when the patient was moved, and the sensor was also prone to falling off. Other problems were that the image moved as the patient breathed and there was no patient outline on the screen.

The final system, that proved successful, is a variation on the original system. A box, containing a single sensor and several buttons, is used to set the static markers. Once the markers are set, another button then instructs the system to set the viewing direction according to the static markers. The software is programmed so that no image can be obtained when the markers have not been set, since such images would be misleading. Although this system requires the markers to be set whenever the patient is moved, having the control buttons on the box means that the procedure can be carried out quickly.

7.2.8 Colonoscope Auto Cut-off Feature

A problem found during the clinical trials was that the image of the colonoscope which was outside the patient, tended to obscure the image of the part inside. This occurs because when the patient is on their left side, the viewing direction is toward the endoscopist and therefore toward the remainder of the colonoscope. To prevent this happening, an auto cut-off feature was introduced. This uses the position of the three static markers to determine how much of the colonoscope is inside the patient. Any part outside the patient is cut-off and is not displayed. Prior to this being introduced, the system required an assistant to operate the keyboard to set the amount of colonoscope that was being displayed.

7.2.9 Saving Images To Disk

At the request of the endoscopists a 'save to disk' feature was incorporated, which saves the screen images to disk at a rate of one image per second. If the actual images as seen on the

screen were saved to disk as bitmaps, this would produce 300 kilobytes of data per second or 1 gigabyte per hour, and this would clearly be unacceptable. Instead, the co-ordinates of each sensor are saved, together with the co-ordinates of the static markers. This reduces the data rate to about 500 bytes per second or 1.8 megabyte per hour. It also has the advantage that the viewing parameters can be changed when the images are viewed rather than when the images are stored.

7.2.10 Block Diagram Of Software

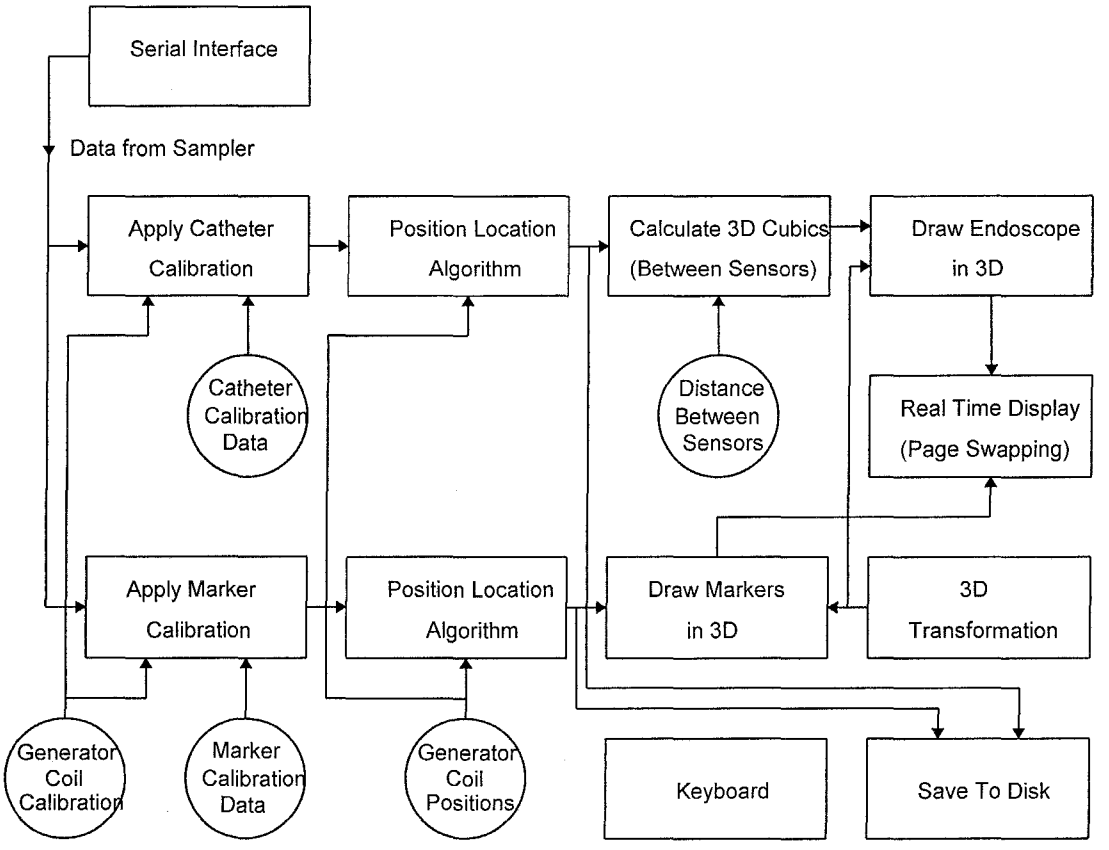


Fig 7.3: Data diagram of the main system unit software

Figure 7.3 shows a data diagram of the system. The boxes represent functional units or modules, the circles represent constants and the wires represent data. The arrows show the direction of data flow. The diagram shows how the data is received from the sampler,

normalised using the calibration data and converted into co-ordinates. The co-ordinates are then shifted, rotated and scaled using the 3D transformation and then drawn on the real time display. The information received by the keyboard module does not enter the data flow diagram directly, but instead issues commands to other modules, such as the 3D transformation module.

7.3 Sampler Software

The sampler software is written directly in assembly language and runs on the Motorola DSP56001 processor. The software takes the sixteen outputs from the A to D converters, multiplies them by a window function stored in ROM and selects the 10 kHz component using a matched filter. The sixteen inphase and quadrature results are then stored until required by the main processor.

The data capture process is initiated by the software in the main system unit. After this, the sampler software controls the sequencing of the drive to the nine generator coils and the data capture and filtering, until all nine sets of data have been stored. The sampler unit then indicates that the data capture process is complete and waits until the main system software is ready for the data. After sending the data the sampler is ready to repeat the process. Each channel has an overload flag which is set when the A to D range is exceeded. This occurs when the sensor is placed too near a field generator, and indicates that the data from that sensor should be discarded.

It is not sufficient that all the A to D's to sample at 40 kHz. It is also necessary that they start the sampling process when the 10 kHz signal, which drives the generator coils, is at a phase of zero. This synchronisation process is controlled by hardware, but is initiated by software. It is also important that the serial outputs from each A to D converter occur in synchronisation, since otherwise the sixteen channels of data would not reach the DSP chip at the same time.

This is achieved by operating one A to D converter in 'master' mode and the remaining ones in 'slave' mode. The A to D converters are also initialised simultaneously when the sampler unit is initialised, using a circuit which is synchronised to the A to D clock. This is recommended by the A to D manufacturers, and ensures that the all the input stages, which oversample the signals, are in phase.

Watchdog software is included in the main unit software to ensure that the sampler is operating correctly. If not, the sampler is reinitialised. The main purpose of this is to prevent the software from locking up if then sampler is disconnected while the system is running. If the sampler is subsequently reconnected, the imager continues.

7.4 References

[7.1] Dennis Harris : "Computer graphics and applications", Chapman and Hall Computing (1984).

[7.2] Steven Harrington : "Computer graphics - a programming approach", McGraw-Hill (1983).

Chapter 8

Laboratory Evaluation Of The Positional Accuracy Of The System

8.1 Method Of Verification Of The Positional Accuracy

Before the imaging system was used in clinical trials, tests were carried out to establish its basic positional accuracy. A single position sensor was attached to a computer controlled scanning frame so that it could be scanned in a plane above the generator coils. The scanning frame was able to position the sensor at known positions and with a known accuracy (± 0.5 mm) so that the positional results from the system could be compared. Since the scanning frame was made from metal, the sensor was mounted on a 1 metre long wooden extension which bolted to the scanning arm.

The tests consisted of a 40 cm by 40 cm xy raster scan with measurements taken at 40 mm intervals, making 121 measurements in total. The scanner was made to stop for each measurement. The z distance (distance between the sensor and the generator coil plane) was kept constant at 0.3 metres throughout the scan.

8.2 Results

Initial results showed that the position system reliably converged to sensible values, with no invalid points or points converging to completely incorrect solutions. However, some distortion was present, which exceeded that expected by the simulations of the algorithm. Figure 8.1 shows the initial results.

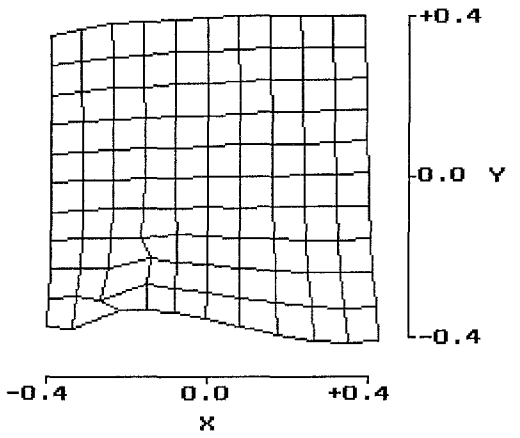


Fig 8.1: Initial results of positional accuracy. z=0.3 metres.

The maximum error and RMS error for this scan are:

Maximum Error	RMS Error
37 mm	16 mm

Fig 8.2: Maximum error and RMS error for the initial results

Many tests and modifications were made to the system, including trying different generator coils and both position algorithms, with no improvement to the scan results. The problem turned out to be two earthing problems in the power amplifiers which drive the generator coils, causing a certain amount of crosstalk between channels. Although only one generator coil is intended to be driven at a given instance, this crosstalk was causing about 2% of drive to be fed to the coils which were supposed to be switched off.

The first earthing problem, responsible for 0.5 % of the crosstalk, was due to the generator coils sharing a common earth connection at the connector to the main system unit. Separate earth wires are used in the cables after the connector.

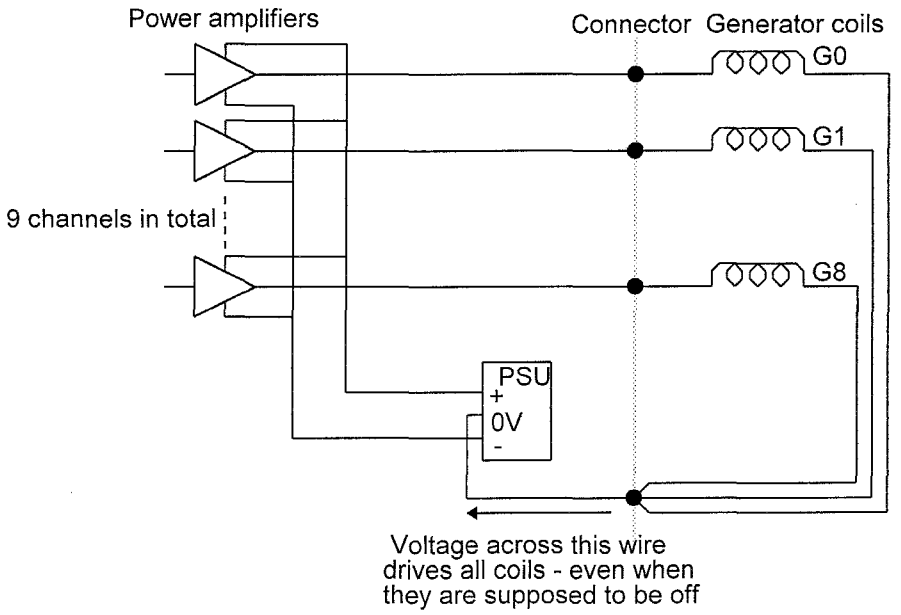


Fig 8.3: Earthing problem with the output stage of the power amplifiers

The problem arises because the outputs of all the amplifiers are referenced to the 0V line from the power supply. When an amplifier is not driving its generator coil, the amplifier output is effectively a short circuit. Any voltage across the wire between the 0V line of the power supply and the 0V connection at the connector, then drives all the generator coils. The best solution to this problem would be to change connector for an 18 way type, giving independent earthing to all channels, but a suitable component could not be found. As a quick solution to this problem, the wire causing the problem was replaced by a number of thick gauge wires, to reduce the voltage across it.

A worse problem, responsible for the majority of the crosstalk, was occurring at the input stage to the power amplifiers. When an amplifier is switched off, its input is shorted to ground by an electronic switch. However, in the original design, this ground point was the same for all nine power amplifiers. Since the power amplifiers are physically in difference places, any 10 kHz present in the earth wiring (mainly in the output stage) enters the input stage of all the power amplifiers.

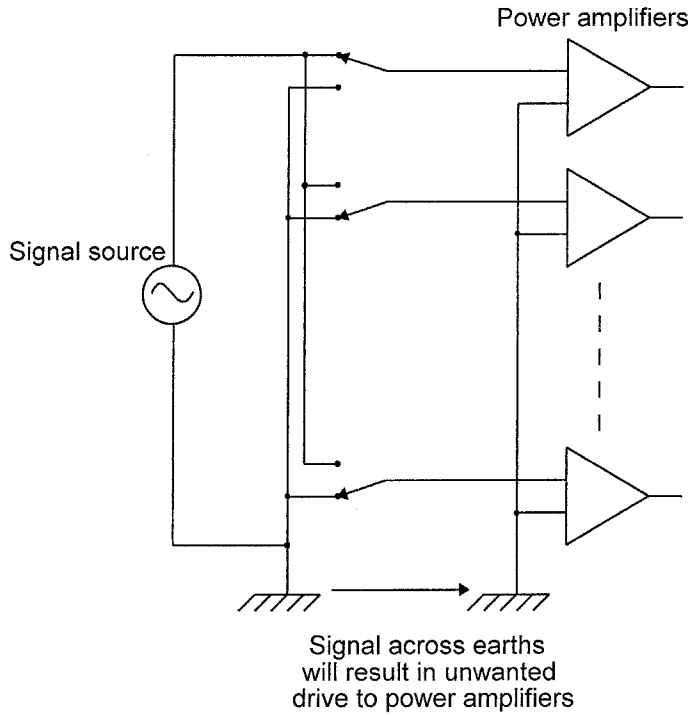


Fig 8.4: Earthing problem with the input stage of the power amplifiers

One way to solve this problem is by leaving the amplifier inputs floating when they are not being driven. This was tried, and was found to improve the situation, but the high impedance lines were now causing crosstalk due to capacitance between tracks in the PCB layout. A solution which worked better was to rewire the switches so that they earthed the amplifier input to the amplifier ground rather than to the local ground near the switches. The seemed the best compromise without having to redesign and rebuild the circuitry. The results are still not perfect (about 0.4% crosstalk), due to the high output currents causing other earthing problems.

A solution which would solve all these problems is to have a single power amplifier followed by a nine way electronic switch, which left the generator coils open circuit when they were not being driven. This option will be used in later versions. It was deliberately rejected before since any capacitance between the adjacent twisted pair wires to the generator coils would cause unwanted crosstalk between the drive to the coils. However, calculating this

capacitance shows that the resulting crosstalk would be orders of magnitudes less than is the case with the current design.

Figure 8.5 shows the results from a scan taken after the earthing arrangement was improved. The accuracy is clearly considerably improved.

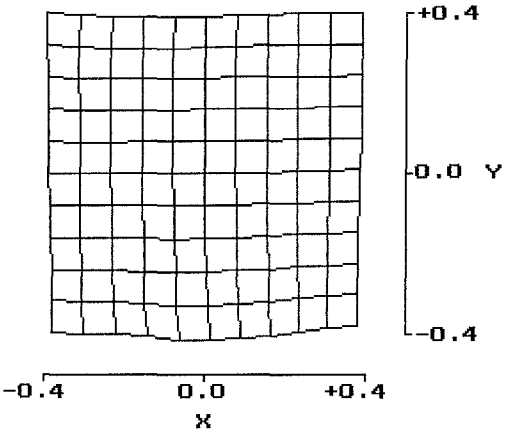


Fig 8.5: Positional accuracy after the earthing was modified.
Z=0.3 metres.

The maximum error and RMS error for this scan are:

Maximum Error	RMS Error
21 mm	7.8 mm

Fig 8.6: Maximum error and RMS error for the system with the modified earthing arrangement

These results are still about an order of magnitude worse than expected by the simulations in chapter 4 (for the given size of generator coil). This is not too serious a problem, since the larger errors only occur at one edge of the range, and at the centre the system is accurate to a few millimetres. The distortion also reduces with increased z.

8.3 Conclusion

The performance of the system is certainly good enough for the application of colonoscopy. The expected accuracy of the algorithm, as predicted by the simulations in chapter 4 are not quite achieved by the current system, probably due to design problems which have not yet been addressed. For colonoscopy, the extra error introduced by the presence of the metal of the colonoscope (up to 20 mm) dictates that it is not worth the effort improving the basic system accuracy. However it would be an interesting future problem to solve.

Chapter 9

Clinical Evaluation

9.1 Preliminary Studies

The clinical evaluation of the imaging system started with two pilot studies on small numbers of patients, to prove that the system worked inside real patients and in the presence of the other medical equipment used during colonoscopy. For the first of these studies, the multiple sensor catheter had not been built and so this study used a single sensor which was withdrawn from the biopsy channel when an image was required. During withdrawal, the system continuously monitored the sensor position, and drew its path on the screen. The results from this study were encouraging and were published in the LANCET [9.1]. During one of these examinations, two x-rays were taken of the patient, one of the front (Anterior-Posterior) view and one of the side (Left-Lateral) view, so that the accuracy of the imager could be compared with the x-ray pictures. These results are shown in fig 9.1.

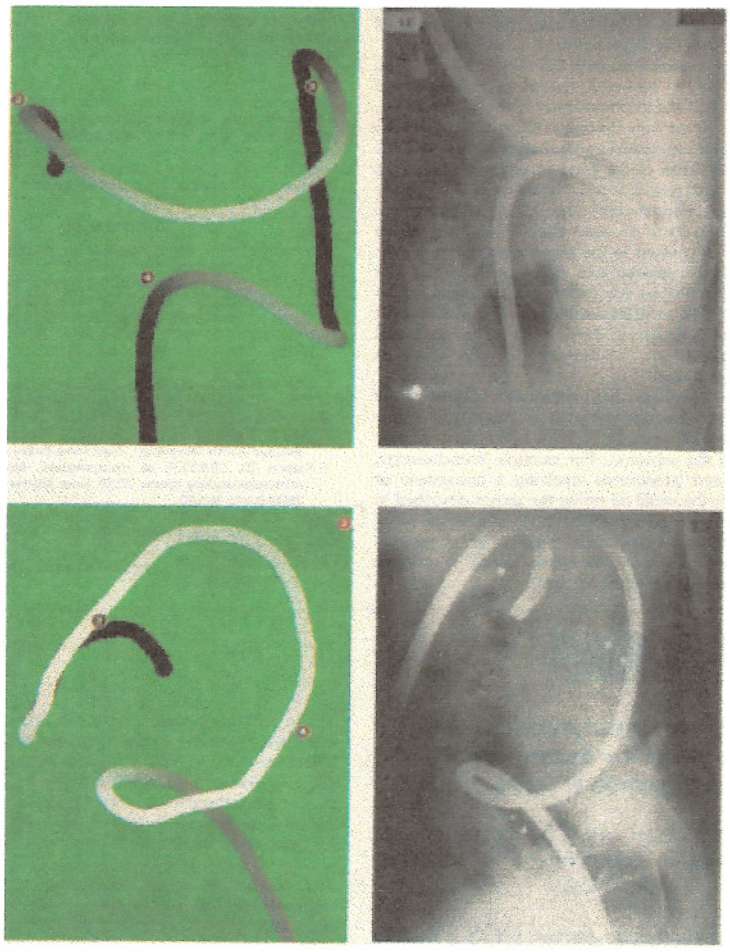


Fig 9.1: Comparison between images obtained using the system and x-ray images.
(a) Front view, (b) Side view

The second pilot study used a multiple sensor catheter, to produce a real time image without having to withdraw the catheter. Due to constructional difficulty, this early version comprised only eight sensors, and was only able to image 70 cm of the colonoscope. Being real time, this system proved much more useful for removing loops during the insertion of the colonoscope. However, it also confirmed that imaging only part of the colonoscope was not sufficient since loops were found to form all the way along it. For the remainder of the studies, a catheter containing fourteen sensors was built, imaging the entire 1.6 metres of the colonoscope. Figure 9.2 shows a typical view of the imager being used. The two imager views (AP and LL) are superimposed on the standard endoscopic view. They are also displayed on a separate computer monitor.

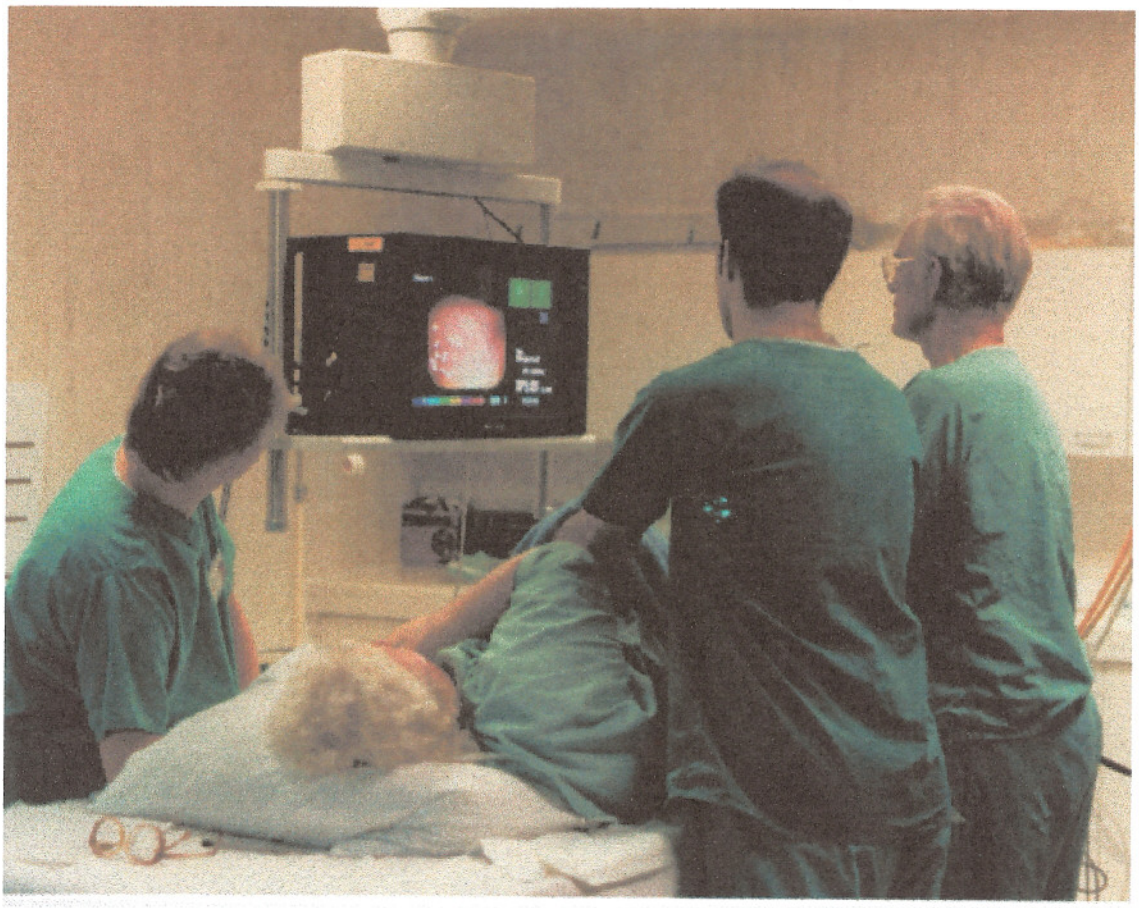


Fig 9.2: View of colonoscopy procedure. The imager views are superimposed on the endoscopic view

9.2 Main Study

The main study, which took place over a six week period, involved using the imager on 61 consecutive patients. Of these, 6 patients were excluded from the study, mainly due to failed bowel preparation. The goal of these tests was to determine how well the system fulfilled its original aims as an aid to colonoscopy. All patients gave informed consent and the study was approved by the local medical ethical committee. The examinations were carried out by one experienced endoscopist with experience of more than 1000 colon examinations.

Although the imager was used with all the patients, the endoscopist and endoscopy assistant were only able to see the imager view for 26 of these. For later study, the images from the imaging system were recorded to disk and the verbal comments made by the endoscopist

during the examination were recorded on paper. In this way the written comments and stored images could be combined after the examination for further study. Patient information, such as age, sex, previous surgery, previous complaints and medication were recorded for each patient.

When the design decision was made to use the biopsy channel for the sensors, it was assumed that it could be blocked throughout intubation. However, during the pilot studies it was found that the suction facility, which is used for deflating the colon, also uses the biopsy channel and therefore did not operate whilst the sensor catheter was inserted. Inflation and deflation of the gut is an important technique in colonoscopy, especially in getting to the caecum at the final stage of intubation. At this point, the loops that inevitably form along the colonoscope prevent any forward motion at the tip despite further insertion by the endoscopist. Suction can then be used to make the colon shrink onto the colonoscope, forcing the tip forward into the caecum. Since suction is only necessary occasionally, the catheter was left in place as much as possible, but was temporarily removed where suction was necessary. In the long term, this problem could be completely eliminated by mounting the sensors inside the endoscope or by making the sensor catheter much smaller. An alternative approach is to use a research colonoscope which has two biopsy channels. This was tried, but the dual channel instrument was an old fibre optic design and proved difficult to use.

To enable the system to produce the required views, i.e. the front or AP view and side or LL view, the patient orientation needs to be known. To enter the patient orientation into the imager system the static marker system was used, as described in chapter 7. Static markers were placed at three standard positions on the patient: (1) at the hepatic flexure, (2) at the splenic flexure and (3) at the anus. The markers are set by holding a sensor probe at the required position and pressing a button corresponding to the marker number. The system measures the probe position and then displays it as a marker on the screen. Once all three markers are set, the software uses them to set the required views. Two further static markers were recorded, one at the middle of the back and one at the umbilicus, to enable the AP

distance to be measured. If the position of the patient was changed during the procedure, which occurred up to four times, then the positions of the markers were set again.

A further benefit of the imager is being able to see where abdominal pressure is being applied. Abdominal pressure means applying pressure to the abdomen to prevent the formation of loops in the colonoscope and is carried out by the endoscopist's assistant. Prior to the existence of the imager, abdominal pressure was a somewhat hit or miss affair, since there was no way of telling whether force was being applied in a suitable place. Thus the potential benefits of this technique were not realised. Using the imager, a sensor can be attached to the hand of the endoscopist's assistant and its position displayed on the screen with the image of the colonoscope. The assistant can then see the effect that the abdominal pressure has on the colonoscope and hence where best to apply it.

9.3 Results

In the study as a whole the use of the imager proved to be highly beneficial. During the examinations where the endoscopist could not see the imager, he was asked where he thought the tip was. On many occasions, his estimate was wrong - always on the side of thinking that the colonoscope was inserted further than it actually was. For example, in one patient a carcinoma which was at the sigmoid / descending junction was incorrectly thought to be at the hepatic flexure. Where surgery is required as a result of colonoscopy, it is important that the estimate is as accurate as possible.

Total colonoscopy, which is insertion of the instrument to the caecum or surgical anastomosis, was achieved in all patients except one. In this case the endoscopist was without the imager view. In this case, the imager cannot be proved to be of significant advantage in achieving total colonoscopy. However, the endoscopist performing the procedures was highly skilled and experienced. Future studies, with less experienced endoscopists who do not

always achieve total colonoscopy, would show how much of an advantage the imager was in improving their performance.

Nearly half of the loops that formed during the tests were incorrectly diagnosed by the endoscopist, leading to incorrect action being taken to remove the loops. A sigmoid spiral loop was correctly diagnosed on 24 out of 28 occasions and a deep transverse loop 3 times out of 5. However, alpha looping occurred 9 times but was only correctly recognised once. Also all 3 reverse sigmoid spiral loops and 1 gamma loop were incorrectly diagnosed.

The AP distance (depth of the abdomen) was found to have an interesting correlation with the difficulty of performing the procedure. Where the AP distance was small, the procedure was generally fairly straight forward. However, when the AP distance was large, it could be expected that the procedure would take some time and the patient experience some discomfort. The problems occur because, with a deep abdomen, the colonoscope loops to the back of the pelvis and then to the front of the abdomen before completing its passage through the sigmoid and heading up the descending colon. These loops result in much of the forward force component being lost, causing the procedure to be difficult. The correlation is only visible with the left lateral view and since when fluoroscopy is used, it is used to obtain an AP view, it has not been observed before.

Abdominal pressure was used on 10 occasions in each of the study groups. When the imager was not visible to the assistant applying the pressure, it was only effective in one case. In fact the stored images showed that the pressure was clearly in the wrong position to provide any benefit. However, when the imager was available, the pressure enabled 7 of the 10 loops to be immediately removed.

One patient had undertaken a colonoscopy examination on two previous occasions, and in neither case was total colonoscopy possible. The endoscopist that carried out one of these procedures reported that 'the instrument was of insufficient length'. The patient, still suffering

from unexplained anaemia was then admitted for a third examination. With the assistance of the imager, a large gamma loop was found to form in the colonoscope as it was fed through the transverse colon. Gamma loops can consume the entire 1.6 m of the instrument. With the help of the imager, the loop was removed and the colon was then examined to the caecum. This patient was found to have a cancer in the caecum and this had not been found by the incomplete examinations carried out without the use of the imager.

One result of particular importance to the design of the imager system was the discovery the metal base had some effect of the accuracy of the positioning system. For reasons discussed earlier, the metal top of the bed was removed and was replaced by a wooden top. The metal base, which contains the raising and lowering mechanism was retained. However, during the tests it was found that the metal base caused positional errors, especially when the bed top was not raised to its highest point. With the bed at its lowest point, the metal base is only 4 inches from the generator coils and causes considerable distortion. The solution to this problem is to use a bed which is entirely non-conductive and non-ferromagnetic such as a carbon fibre bed (as used for computed tomography).

9.4 Images Obtained Using The system

The images in fig 9.3 were obtained during the course of the clinical trials, and show four loops that commonly occur during colonoscopy. Standard colonoscopy text books describe these loops in great detail, together with idea on how to detect the presence of a given loop and how to remove it. However, it is still largely experience and guesswork.

Figure 9.4 shows the procedure for straightening an alpha loop, using a combination of pulling back and clockwise twist. With the imager, this procedure is intuitive since the behaviour of the colonoscope can be seen throughout the manoeuvre. Without the imager, it is much more difficult, since the endoscopist will not even be sure of the presence of the

alpha loop when he starts the manoeuvre. The procedure must therefore be carried out assuming this type of loop is present, and then only after its completion will it be clearer whether the guess was correct or not.

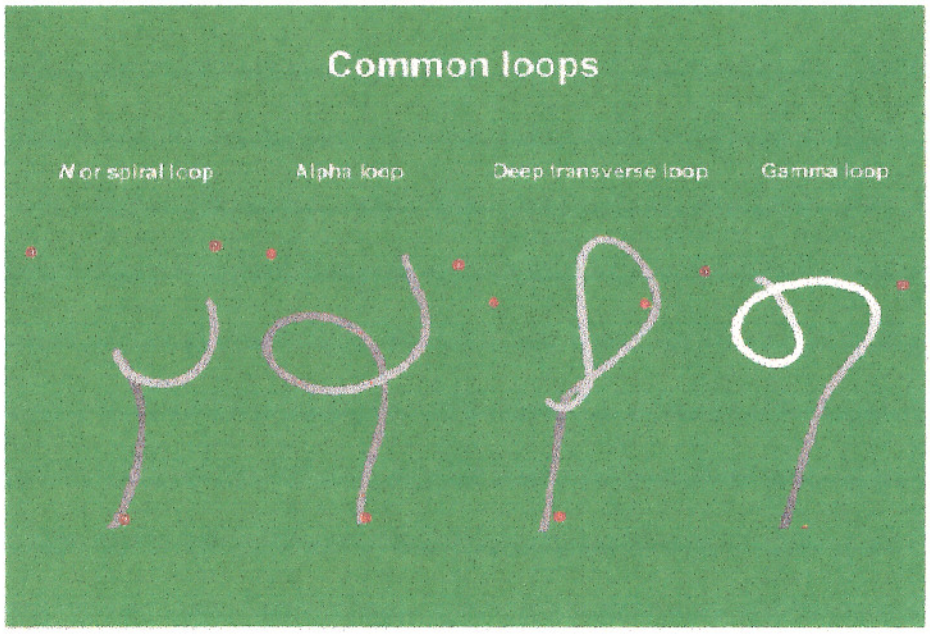


Fig 9.3: Common loops that form during colonoscopy, obtained using the imager

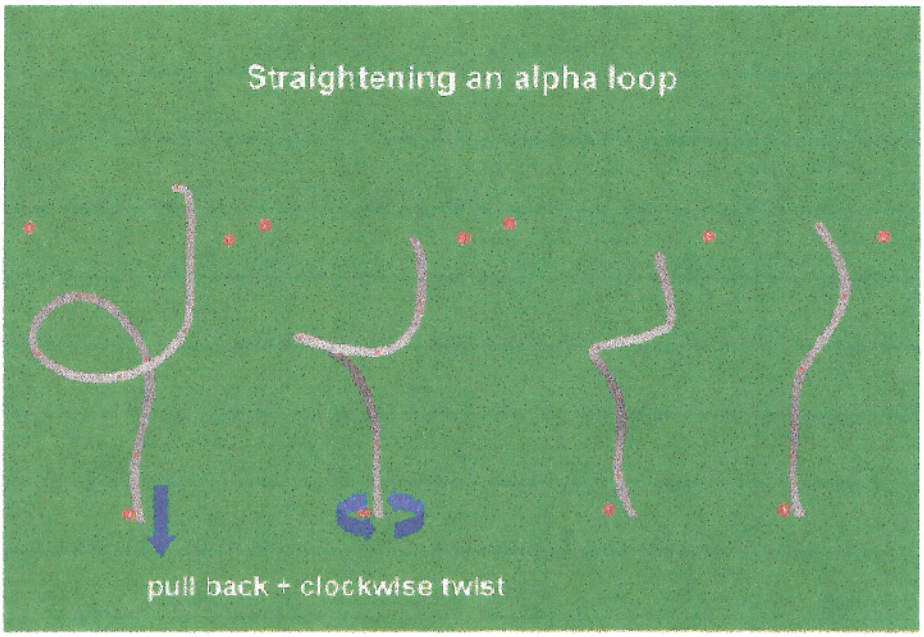


Fig 9.4: Straightening an alpha loop

Figure 9.5 shows abdominal pressure being applied, with a sensor attached to the hand of the assistant applying the pressure. The sensor is shown in blue. It is quite clear, in this case, that the abdominal pressure has reduced the sigmoid looping. The imager will also show when abdominal pressure is not helpful, for example where the abdominal wall is particularly thick.

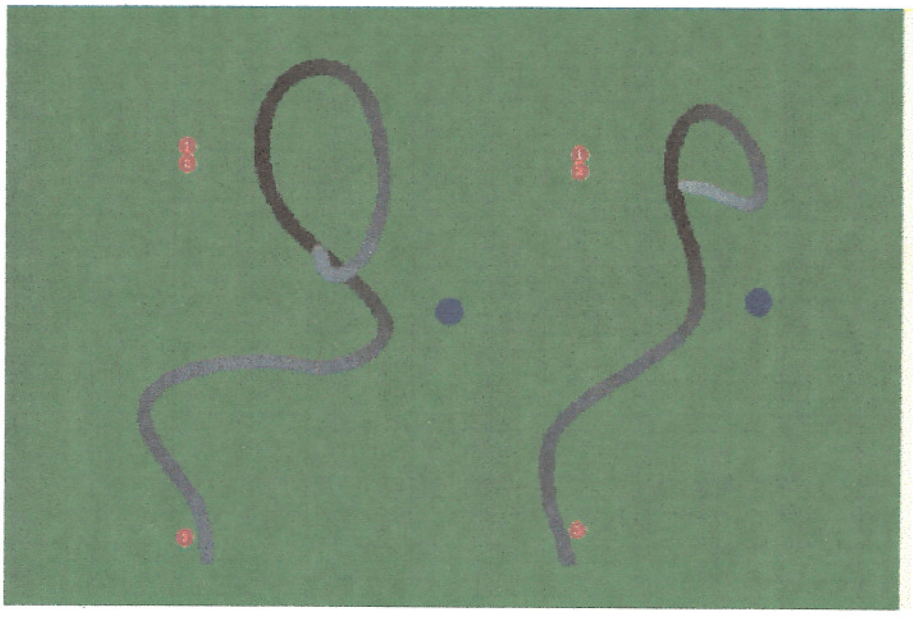


Fig 9.5: Left lateral view showing abdominal pressure correctly applied to reduce looping. The hand sensor is shown in blue. The left hand image shows the colonoscope before pressure is applied

9.5 Conclusion

Prior to the use of the imager, the knowledge of when to use clockwise or anticlockwise twist together with pushing and pulling of the colonoscope, could only be gained by significant experience. Even with that experience, finding the correct manoeuvre to resolve a particular problem requires a good deal of experimentation. The imager enables the endoscopist to see exactly what he is doing and how the colonoscope is responding, thus enabling the correct manoeuvre to be found almost instantly.

The medical trials discussed here were carried out by a highly experienced endoscopist who was trained by a colonoscopist of international reputation. The benefits are expected to be

even greater for a less experienced person. Future clinical studies will be carried out with trainee endoscopists to establish whether the imager will reduce the training time and ease the learning curve. Studies will also be carried out to determine how much quicker an examination can be carried out when the imager is used. This fact is difficult to extract from the data in this chapter, since when using the imager the endoscopist tended to get carried away trying all sorts of manoeuvres to see their effect!

9.6 References

[9.1] J.S. Bladen, A.P. Anderson, G.D. Bell, B. Rameh and D.J.T Heatley : "Non-radiological technique for three-dimensional imaging of endoscopes", The LANCET, Vol. 341, pp 719-722, 20th March 1993.

Chapter 10

Software Tools Used In The Design Of The Endoscope Location System

10.1 Introduction

In addition to the main system software, many different programs were written during the course of the development of the system to help design and test various aspects of it, including field calculation, position algorithm and signal processing. Three of these programs are written as interactive Microsoft Windows™ applications. The first of these, MAGCALC, enables the user to modify all aspects of the endoscope imaging system, such as generator coil size or system frequency, and then simulate the positional accuracy that such a system would produce. This program also enables the effects of a metal frame around the bed to be simulated. SENSOR3D is a 3D finite difference program, optimised to calculate the effects of various core materials inside the sensor and enables the optimum size and shape of sensor to be designed. ENDOVIEW is written for use by clinicians, and enables complete patient examinations to be recalled after the examination has been completed. The program is written to help analyse the image data, and provides tools for making measurements of loop size, bend angles and radius of curvature of bends.

10.2 MAGCALC

The purpose of being able to simulate the behaviour of the endoscope position location system, is to help optimise and debug the system, especially when a single effect, in this case positional accuracy, may be caused by many variables. In that situation an empirical approach is not helpful. For example, the first prototype performed rather badly due to small amounts

of crosstalk between the power amplifiers and the wires to the sensors. Without access to a simulation, the errors were assumed to be due to the finite size of the generator coils, rather than the crosstalk, which appeared insignificant compared to the signal strength. The use of a simulator showed the significance of the crosstalk, but also showed that the generator coils could be made significantly larger without loss of accuracy.

Various commercial software packages, using both finite difference and finite element solutions, were considered for analysing the endoscope location system. All these packages suffered from extreme high cost and inflexibility of use with regard to modifying the system design. Being non-symmetrical, the 3D situation has to be built up using multiple 2D layers and this can take some weeks or months to enter for a single design. The problem is made worse as high mesh spatial resolution is required near to the generator and sensor coils, but cannot be used throughout the volume due to memory restrictions. Using a working volume of 2 m x 2 m x 2 m and a linear resolution of 10 mm typically requires 256 MB of memory and a correspondingly long computation time. However, using a variable resolution mesh makes it difficult to modify the design, for example by moving a generator coil assembly, without starting from scratch each time. The large mesh has to be used because the boundary conditions are not known and therefore have to be made distant from the generator and sensor coils. Once the field results are obtained, they have to be ported to another program which performs the position location algorithm in order to obtain the estimate of sensor position. A further problem is that many packages are unable to cope with the large air gap between the generator coil and the sensor coil and suffer from large internal rounding errors and premature convergence conditions.

This process is cumbersome and also error prone since manual intervention is required to extract the relevant field calculations. It is also so lengthy that proper evaluation of the system with regard to all possible geometries and signal conditions is not possible. To remedy this situation, a special purpose field calculation package was designed which is optimised for calculating the performance of the complete endoscope location system. This package, called

MAGCALC, enables all design parameters, such as the position and size of the generator coils, to be altered and then displays a wide range of results, including field strengths and calculated positional accuracy. In also calculates the range and standard deviation of the position of a single static sensor in the presence of signal noise.

MAGCALC uses neither finite difference nor finite element analysis, but instead uses elementary magnetic field methods and circuital laws to calculate the field produced by each current carrying element of the system. The Biot-Savart law is used to calculate the field at the sensor due to the currents in the generator coil assemblies and Faradays law of induction is used to work out the induced voltage in the sensor. Although it is difficult to write simple analytical expressions for the fields produced by the generator coils, it is possible to write partial expressions which can be summed vectorially in the computer to produce the final field result. This process is described in Chapter 4.

Two causes of field distortion are eddy currents in the colonoscope and eddy currents in the bed. The conventional approach to solving eddy current problems is derived as follows:

The induced emf is given by equation 10.1 (**A** is the magnetic vector potential).

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t} \tag{10.1}$$

The current density is given by:

$$\mathbf{J} = \sigma \mathbf{E} \tag{10.2}$$

$$\mathbf{J} = -\sigma \frac{\partial \mathbf{A}}{\partial t} \tag{10.3}$$

Substituting into Maxwells equations:

$$\mathbf{B} = \mu \mathbf{H} \tag{10.4}$$

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{10.5}$$

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{10.6}$$

Therefore,

$$\nabla \times \frac{1}{\mu} \nabla \times \mathbf{A} + \sigma \frac{\partial \mathbf{A}}{\partial t} = 0 \tag{10.7}$$

This equation can be solved for simple symmetrical geometries [10.1, 10.2, 10.3, 10.4] but is much more difficult for the present situation, either for calculating the eddy currents in the bed or the colonoscope.

An alternative approach to calculating the effect of a metallic bed, is to treat it as a conductive metal loop, with a known resistance and inductance. The field from the bed is calculated by summing the induced voltage in elements around the bed frame to produce the induced voltage in the loop, and then calculating the field produced by the current that would flow if such a voltage source is inserted at a point in the loop. The field thus calculated is added to the field that comes directly from the generator coils. This is the approach that is used by MAGCALC.

The calculation of eddy currents in the colonoscope is a difficult problem and has not been attempted. It would require a detailed knowledge of the construction of the colonoscope, including the types and conductivities of the materials used. The results would not necessarily be useful, since practical results adequately show the distortion produced by the colonoscope. What is more useful, is to understand general trends, such as whether a higher or lower frequency would reduce the errors. This is discussed in chapter 6.

The results from MAGCALC are three dimensional and may be viewed in either the x, y or z directions. A plan view (z view) is shown in figure 10.1.

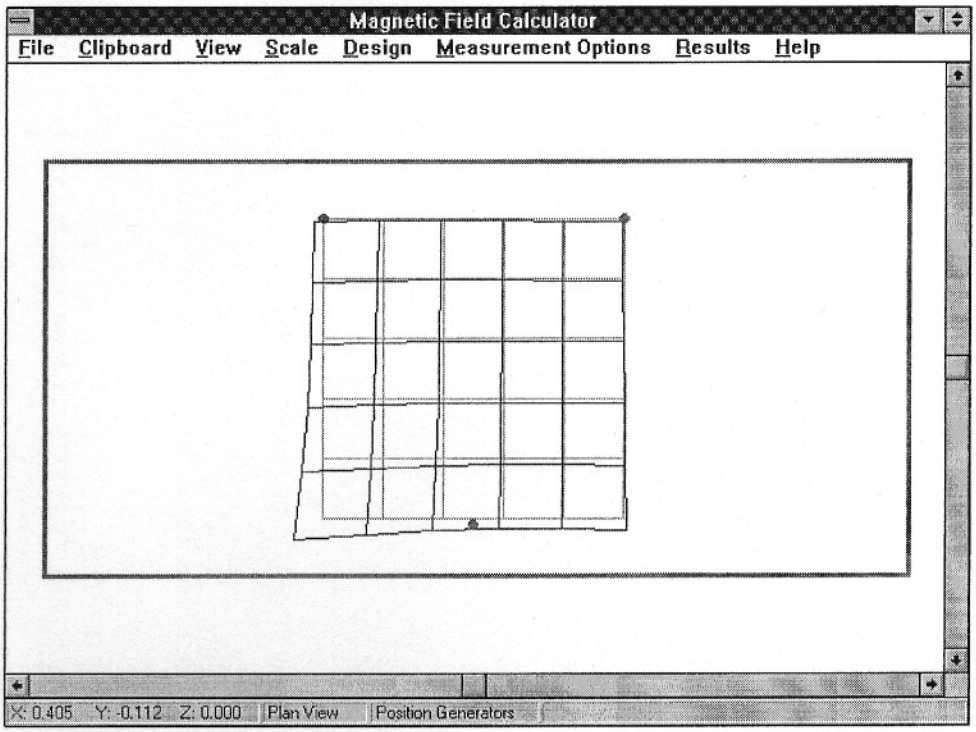


Fig 10.1: A view of MAGCALC showing the predicted positional accuracy for a sensor offset of 50 μV

10.3 SENSOR3D

The program SENSOR3D was written as a companion to MAGCALC, and enables the effects of different core materials and sensor shapes to be predicted. The program uses the established 3D finite difference technique to calculate the induced voltage in the sensor winding in the presence of a magnetic field. The magnetic field is established by setting the boundary conditions.

The problem is made straight forward by the fact that there are no field sources near to the sensor. This means that the problem, although using magnetic fields, is identical to solving the problem of electric currents through a conductive volume.

The field behaviour is given by :

$$\nabla \cdot (\mu \nabla \psi) = 0 \tag{10.8}$$

The derivation of this is given is Section 6.5. μ represents the material permeability distribution and ψ the magnetic scalar potential. This is analogous to the electric field case:

$$\nabla \cdot \left(\frac{1}{\sigma} \nabla \phi \right) = 0 \tag{10.9}$$

where σ is the conductivity and ϕ is the electric field potential.

This problem can be solved by standard field packages since it is a more straightforward situation that simulating the whole system. However, for reasons of speed and consistency it was felt appropriate to write a purpose designed program. The program takes the sensor size, the number of turns on the sensor, the calculation volume and the boundary conditions and then calculates the induced EMF in the sensor.

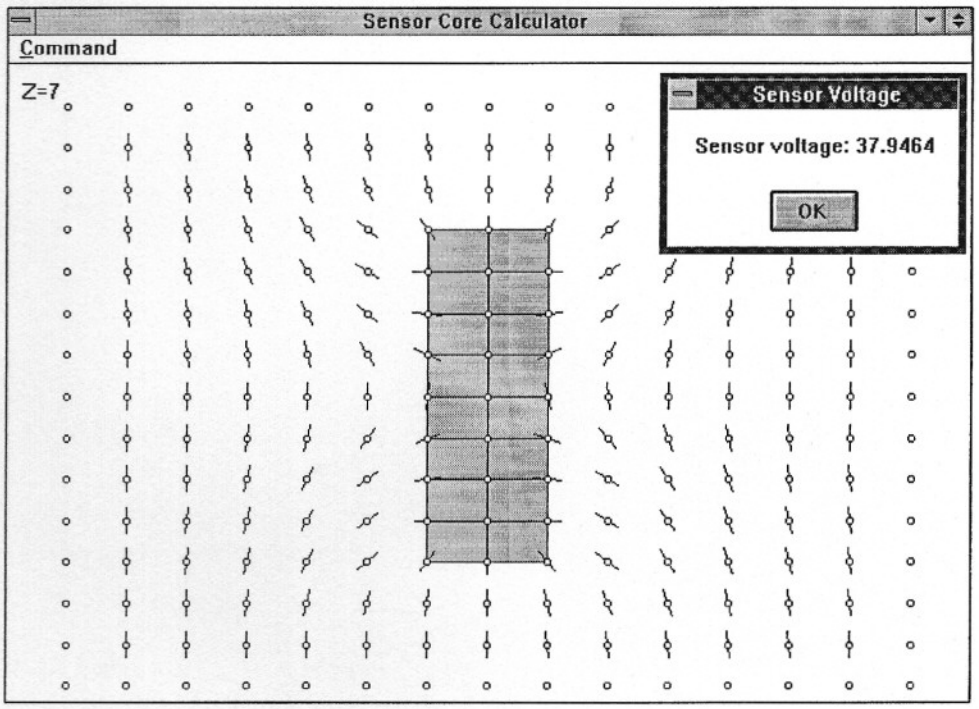


Fig 10.2: A typical view of SENSOR3D

10.4 ENDOVIEW

Whilst the endoscope imaging system is in use, a screen shot is saved to hard disk at a rate of one image per second. To reduce the data rate, and also to enable 3D manipulation of the data, the sensor positional data is stored, rather that a bitmap of the screen. The time is also recorded with each image to enable the intubation time to be established. Figure 10.3 shows a typical view taken from real patient data.

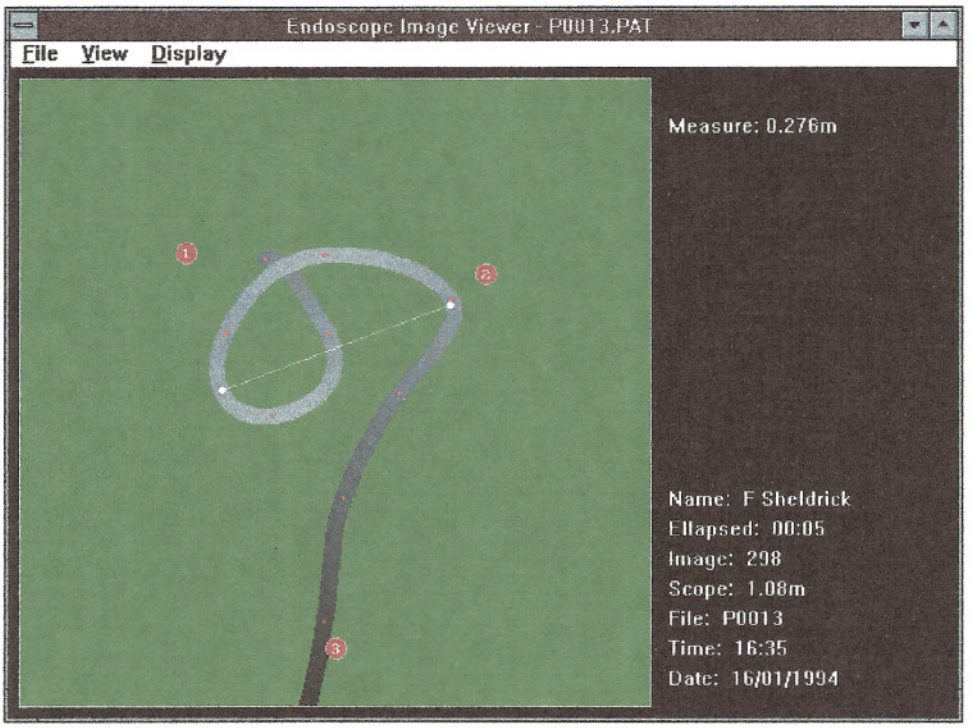


Fig 10.3: ENDOVIEW - a program for viewing the images obtained during a colonoscopy. Image shows a gamma loop forming in the transverse colon. The red dots on the endoscope indicate the position of the sensors. The white line across the image is a 3D ruler.

ENDOVIEW includes facilities for scaling and rotating the images in 3D, in a similar way to ENDOSCAN, the program which drives the imaging system.

The controls are as follows:

Key	Function
A	Use markers to set an Anterior-Posterior View
L	Use markers to set an Left Lateral View
X	Move image right
Shift-X	Move image left
Y	Move image up
Shift-Y	Move image down
Z	Move image toward viewer
Shift-Z	Move image away from viewer
Left Arrow	Rotate image left
Right Arrow	Rotate image right
Up Arrow	Rotate image up
Down Arrow	Rotate image down
Delete	Rotate image anti-clockwise
End	Rotate image clockwise
Page Up	Show previous image
Page Down	Show next image

Fig 10.4: Keys used in ENDOVIEW

A 3D measurement facility is also included which enables distances to be measured and is selected from the VIEW menu. This comprises a ruler whose ends may be positioned using the mouse. Since the ruler is in 3D space and the display is only 2D, it must be positioned in both the AP view and the LL view before the displayed distance is valid.

10.5 References

[10.1] G.A. Girgus and A. Bastawros : "Nondestructive eddy current testing for the measurement of conductivity and surface buckling of metallic sheets", IEEE Transactions on Instrumentation and Measurement, Volume Im-35 Number 4, December 1986.

[10.2] M. Koizumi and M. Onizawa : "Computational method of three dimensional eddy current by using volume integral equation method", IEEE Transactions on Magnetics, Volume 27 Number 5, September 1991.

[10.3] A. Nicolas : "A boundary integral approach for eddy current calculation", IEEE Transactions on Magnetics, Volume MAG-19 Number 6, November 1983.

[10.4] G. Rubinacci : "Numerical computation of the eddy currents on the vacuum vessel of a tokamak", IEEE Transactions on Magnetics, Volume MAG-19 Number 6, November 1983.

Chapter 11

Conclusion

A novel positioning and imaging system based on magnetic field location has been devised and developed. It appears to be unique in that only one vector component of the field is measured enabling a single minute coil to be employed. It is therefore appropriate for medical investigations where the sensors can be inserted into catheters. The system described in this thesis has now been successfully in use for colonoscopy for over a year. It is intended that the equipment is introduced into a number of teaching hospitals throughout the UK where it can be evaluated more extensively.

The imaging system has clearly fulfilled the majority of the goals set in Chapter 1. The system produces a real time 3D image of the entire colonoscope to a worst case accuracy of ±10 mm, despite the effects on the magnetic field by the metal in the colonoscope. In principle, an accuracy of ±1 mm might be expected in the absence of metallic interference. The system operates with any existing commercial colonoscope, and the low frequency, low intensity magnetic fields are considered inherently safe such that it can be used throughout the colonoscopic procedure. The only disadvantage with the current design is that it restricts the amount of suction available during intubation. This problem could be solved by either making the sensor catheter smaller, incorporating it inside the colonoscope or making a sensor array which fits on the outside the colonoscope, such work being best undertaken in conjunction with a colonoscope manufacturing company.

The three consultant endoscopists who have used the system so far are all highly experienced, and even so have been surprised by the amount of useful information the system can offer. The system has enabled patients to be examined where hitherto the procedure might have been too dangerous. The imager also helps to reinforce theories about the behaviour of the

colonoscope, which has not previously been possible. Future studies, involving less experienced colonoscopists, would show the effectiveness of the system as a training tool. Other studies could determine whether having the system significantly reduces the time for colonoscopy and whether it makes the procedure safer.

One fact that became clear during the clinical trials is the inadequacy of current colonoscopes. Before the imager, there was no way of observing how a colonoscope behaves in three dimensions when it is actually inside a patient. With the imager, it was observed that the design of the colonoscope made it difficult to remove some loops even when they could be clearly seen. This perhaps challenges the whole concept of pushing a flexible tube into the colon, and maybe an alternative design of colonoscope based on a different principle is possible.

One interesting application of the system is in conjunction with the British Telecom 'Telelink' system, also known as 'remote surgery'. This involves having a ISDN telephone link (64 kBit/sec) between the endoscopy suite and an expert endoscopist situated at a remote site. The idea is to enable less experienced doctors to carry out specialised operations under the supervision of an expert doctor based at a different site. The telephone link transfers moving pictures which have been specially compressed to match the limited bandwidth of the telephone lines. Before the imager system was in use, the Telelink system was not particularly suitable to colonoscopy, since the remote doctor does not have the tactile feedback that he would have if he was carrying out the procedure himself. With the imager, the images of the colonoscope configuration can be combined with the standard endoscopic view and sent to the remote endoscopist, who can use them to offer instruction to the doctor carrying out the procedure. This system has been tested on three patients and has proved entirely successful.

This imaging technology clearly has other applications in medicine. Capsules containing a position sensor together could be swallowed, to study transit through the alimentary canal without using X-rays. Insertion of Swan-Ganz catheters into the heart could be monitored,

simplifying the procedure which is currently done blind. Naso-gastric tubes (tubes inserted into the stomach via the nose) could be visualised, preventing the tube being inadvertently fed into the lungs, a problem which does occur and with fatal consequences. CT scans are often taken before keyhole surgery is performed, but are not directly available during the procedure. The magnetic field positioning technology could enable the position of the tools used by the surgeon to be mapped directly onto the CT scan data, significantly improving the usefulness of the data.

Whilst the imaging system was being developed, other aids to colonoscopy have been considered by other research groups. One of these is the colonoscopy simulator. This consists of a dummy colonoscope which is inserted into a control box. The control box models the behaviour of a patient by applying a variable force to the colonoscope. Computer software calculates the configuration of the colonoscope that would occur if the colonoscope was inside a real patient and then uses that information to display a realistic image of the inside of the colon. The simulator approach is designed to teach the basics of colonoscopy before the endoscopist attempts the procedure on real patients. Together the simulator and the imager system described in this thesis should ease the learning curve of colonoscopy.

There are non-medical applications for the technology which might be considered. One application that British Telecom is currently considering, is to enable deaf and dumb people to communicate across telecommunication links. Sensors would be mounted in a pair of gloves and would determine the positions of the fingers and hands to enable the sign language to be read by a computer. Another application is as an alternative to the computer mouse. The sensors could detect the hand positions as they interact with objects on the screen, giving a more intuitive user interface.