The application of optical diagnostics to high energy electromagnetic acoustic transducers

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Additional Information:

- A Doctoral Thesis. Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy of Loughborough University.

Metadata Record: https://dspace.lboro.ac.uk/2134/11239

Publisher: © M.T. Carnell

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to: http://creativecommons.org/licenses/by-nc-nd/2.5/
THE APPLICATION OF OPTICAL DIAGNOSTICS TO HIGH ENERGY ELECTROMAGNETIC ACOUSTIC TRANSDUCERS

by

Mark Thomas Carnell

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of Doctor of Philosophy at the Loughborough University of Technology

30th September 1995

© by M T Carnell 1995
ACKNOWLEDGEMENTS

This thesis marks the high point of a desire I have had since a school boy, to understand a little more about the things in the world that surround me, so many of which are a testament to the power of the engineer and the physicist. The partial fulfilment of this goal has been in most part due to the help and support of my family and the friends I have met in my many years of education. I would therefore like to take this opportunity to thank them all, especially my supervisor Dr David Emmony, who during the last 6 years has not only given me the chance to study for a degree, but also guided me through my PhD, a fact for which I will forever be indebted. I would also like to thank the members of the physics department who have made these years of study at Loughborough enjoyable and all the more difficult to leave. A special thanks to John Oakley and the rest of the team in the workshop for their assistance in building the equipment. Thanks are also due to the National Physical Laboratory for the loan of the schlieren lens and to Dr Tom Craig of Huntsman Chemical Co for the polystyrene used in the manufacture of the acoustic lenses.
To Mum, Dad and Nanna.
ABSTRACT

This thesis is concerned with the design and construction of an electromagnetic acoustic transducer (EMAT) and the characterisation of its acoustic field both conventionally, using a hydrophone and with high resolution laser illuminated schlieren techniques. During the early 80s the introduction of the EMAT along with the other types of shock wave source used for lithotripsy, revolutionised the treatment of stone disease. The process of shock wave induced destruction of calculi and the use of shock waves in other areas of medicine will be discussed, along with the causes and effects of stone disease in man. For the first time high temporal and spatial resolution schlieren images of the shock waves and their interaction with simulation kidney stones have been recorded. The technique provides a clearer picture of the fragmentation process and may assist research into the suitability of shock wave treatment in other areas of medicine currently under investigation.

Schlieren studies of the acoustic field have shown the complex structure of not only the EMAT shock wave, but also that associated with cavitation in the field. The primary source of cavitation is due to the rupture and subsequent collapse of bubbles generated in the water by the strong rarefaction phase of the shock wave. The images give evidence for the interaction of these 'primary' cavitation shocks with bubbles in the field, the collapse of some of these bubbles giving rise to additional or 'secondary' cavitation shocks. An optical lensing effect introduced by the shock has also been investigated. Objects seen through or immersed in the field of an EMAT shock wave such as cavitation, appear highly distorted, due to the strong positive and negative lensing effects associated with the changing refractive index of the compression and rarefaction cycles of the shock wave.
CONTENTS

Acknowledgements iii
Abstract v
Glossary of medical terms x

CHAPTER 1: Introduction 1

1.1 Introduction

References 4

CHAPTER 2: The History of Stone Disease 6

2.1 Introduction
2.2 Stone disease 8
2.3 The urinary system 10
  2.31 The kidneys 11
  2.32 The renal calculus or kidney stone 13
  2.33 The vesicle calculus or bladder stone 17
  2.34 Early operative procedures 18
    2.341 Perineal lithotomy 18
    2.342 Trans urethal lithotrity 21
    2.343 Suprapubic lithotomy 23
2.4 The Biliary system 24
  2.41 The biliary calculus or gallstone 26
  2.42 Diagnosis 28
  2.43 Treatment alternatives 29
    2.431 Percutaneous cholecystolithotomy 30
    2.432 Methl tert-butyl ether lavage 31
    2.433 Drug (per oral) chemolysis 32
    2.434 Shockwave lithotripsy plus drug chemolysis 32
  2.44 Stone recurrence 33
2.5 Lithotripsy of renal and biliary calculi 34
  2.51 Shockwave lithotripsy 36
    2.511 The electrohydraulic source 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.512 The laser source</td>
<td>37</td>
</tr>
<tr>
<td>2.513 The piezoelectric source</td>
<td>38</td>
</tr>
<tr>
<td>2.514 The electromagnetic source</td>
<td>39</td>
</tr>
<tr>
<td>2.52 Localisation systems</td>
<td>40</td>
</tr>
<tr>
<td>2.53 Stone fragility</td>
<td>41</td>
</tr>
<tr>
<td>2.54 Mechanisms of stone fragmentation</td>
<td>42</td>
</tr>
<tr>
<td>2.541 Cavitation</td>
<td>45</td>
</tr>
<tr>
<td>2.542 Biological effects</td>
<td>46</td>
</tr>
<tr>
<td>2.55 What makes a lithotripter efficient in lithotripsy</td>
<td>47</td>
</tr>
<tr>
<td>2.56 Lithotripter technology and future prospects</td>
<td>49</td>
</tr>
</tbody>
</table>

**References**

**CHAPTER 3: The Generation and Focusing of Acoustic Shock Waves**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>66</td>
</tr>
<tr>
<td>3.2 Acoustic theory</td>
<td>67</td>
</tr>
<tr>
<td>3.21 Acoustic radiation</td>
<td>67</td>
</tr>
<tr>
<td>3.22 Acoustic velocity</td>
<td>68</td>
</tr>
<tr>
<td>3.23 The shock wave</td>
<td>70</td>
</tr>
<tr>
<td>3.24 Specific acoustic impedance</td>
<td>72</td>
</tr>
<tr>
<td>3.25 Reflection and transmission of acoustic waves</td>
<td>73</td>
</tr>
<tr>
<td>3.3 The electromagnetic transducer I operation</td>
<td>75</td>
</tr>
<tr>
<td>3.31 The capacitor discharge system</td>
<td>77</td>
</tr>
<tr>
<td>3.311 The spark gap</td>
<td>79</td>
</tr>
<tr>
<td>3.312 Charging cycle</td>
<td>79</td>
</tr>
<tr>
<td>3.313 Discharge cycle</td>
<td>81</td>
</tr>
<tr>
<td>3.32 The EMAT sound field</td>
<td>82</td>
</tr>
<tr>
<td>3.33 Estimation of shock wave pressure and energy</td>
<td>85</td>
</tr>
<tr>
<td>3.4 Acoustic focusing</td>
<td>87</td>
</tr>
<tr>
<td>3.41 Lens design theory</td>
<td>88</td>
</tr>
<tr>
<td>3.411 Aberrations</td>
<td>89</td>
</tr>
<tr>
<td>3.412 Ray Tracing</td>
<td>91</td>
</tr>
<tr>
<td>3.42 The focal region</td>
<td>93</td>
</tr>
<tr>
<td>3.5 Pressure measurement</td>
<td>94</td>
</tr>
<tr>
<td>3.51 Piezoelectricity</td>
<td>95</td>
</tr>
<tr>
<td>3.52 Hydrophone design</td>
<td>97</td>
</tr>
<tr>
<td>3.521 Prototype hydrophone designs</td>
<td>98</td>
</tr>
</tbody>
</table>
CHAPTER 4: Schlieren Methods

4.1 Introduction 123
4.2 The relationship between refractive index, density and pressure 124
4.3 The Toepler schlieren method 127
  4.31 The focused shadow method 129
  4.32 The scale method 131
4.4 Quantitative analysis using knife edge systems 133
  4.41 The photographic process 135
  4.42 Sensitivity 137
4.5 Optical considerations 140
  4.51 Diffraction 141
4.6 Experimental apparatus 141
  4.61 System layout 142
    4.611 Laser illumination 144
    4.612 Apparatus set up procedure 145
    4.613 Synchronisation and timing 147
  4.62 The precision X-Y translation stage 149
    4.621 The computer interface 150
    4.622 Stepper motor hardware 151
    4.623 Control software 154

References 154

CHAPTER 5: Schlieren Studies

5.1 Experimental method 157
5.2 Schlieren and shadow studies 159
5.3 Quantitative studies 169
  5.31 The assessment of pressure using a knife edge 169
  5.32 The assessment of pressure using the scale method 172
  5.33 Discussion 176
5.4 Shock wave interaction studies 178
  5.41 Interaction with a brass disk 178
  5.42 Interaction with a simulation stone 180
  5.43 Discussion 184
5.5 Cavitation distortion phenomena 186
  5.51 Discussion 191

References 191

CHAPTER 6: Conclusion and Suggestions for Further Work

  6.1 Conclusion 194
  6.2 Suggestions for further work 198

APPENDIX

  I Evaluation of the shock wave intensity 199
  II Sign convention for thick lenses 200
  III Ray deflection due to a refractive index gradient 201
  IV Publication list 204
| **Acidosis:** | An over acid condition of the body fluids. |
| **Astringent:** | A drug to check bleeding. |
| **Bile:** | A fluid secreted by the liver that aids digestion. |
| **Biliary system:** | Composed of the gallbladder, hepatic and common bile ducts. |
| **Bilirubin:** | Principle bile pigment. |
| **Calculus:** | A stone or concretion of material formed within the body. |
| **Cannula:** | A hollow rod. |
| **Chemolysis:** | The use of drugs to dissolve cholesterol stones. |
| **Cholecystectomy:** | The surgical removal of the gallbladder. |
| **Cholecystitis:** | Inflammation of the gallbladder. |
| **Cholecystolithotomy:** | Making an incision into the gallbladder. |
| **Cholesterol:** | A sterol found in most body tissues. |
| **Cirrhosis:** | Chronic disease of the liver. |
| **Colic:** | A severe spasmodic abdominal pain. |
| **Colon:** | The lower and greater part of the intestine. |
| **Congenital:** | A disease or defect existing from birth. |
| **Cystitis:** | Inflammation of the urinary bladder. |
| **Cytotoxic:** | Toxic to cells. |
| **Diuretics:** | Drugs causing an increased output of urine. |
| **Duodenum:** | The first part of the small intestine. |
| **Dysuria:** | Painful or difficult urination. |
| **Endoscope:** | An instrument for viewing the internal parts of the body. |
| **Enema:** | Injection of gas or liquid into the rectum to expel its contents. |
| **Extracorporeal:** | From outside of the body. |
| **Fatty acids:** | Organic compounds consisting of a hydrocarbon chain. |
| **Fistula:** | An abnormal passage between a hollow organ and the body surface. |
| **Glycoprotein:** | A group of compounds consisting of proteins and carbohydrates. |
| **Glycosuria:** | The presence of sugar in the urine. |
| **Gout:** | A disease with inflammation of the small joints. |
| **Haematomas:** | Swelling of clotted blood within the tissue. |
| **Haematuria:** | The presence of blood in the urine. |
Hypercalcuria: Excessive amounts of calcium in the blood.
Hypertension: Increase in blood pressure.
Jaundice: A condition with yellowing of the skin or whites of the eyes.
Lecithin: One of a group of phospholipids.
Lesions: A morbid change in the functioning of an organ.
Lipid: An organic compound i.e. fatty acid, oils etc.
Litholapaxy: A procedure in which stone are crushed and then removed.
Lithotomists: A person undertaking the removal of bladder stones.
Lithotomy: The surgical removal of a stone from the urinary tract.
Lithotripsy: The breaking of stones using ultrasound.
Lithotripsy: The crushing of stones per urethra.
Lithotryptics: The treatment of stone disease using stone dissolving chemicals.
Nephrolithiasis: Kidney stone disease.
Nephrolithotomy: A treatment in which stones are surgically removed from the kidney.
Oedema: Excess of watery fluid in the body.
Pancreatitis: Disease of the pancreas.
Parathyroid gland: Regulates calcium levels in the body.
Parenchymal: The essential or functioning elements of an organ.
Percutaneous: The administration of a remedy through the skin.
Perineum: Region of the body between the anus and the scrotum or vulva.
Peritoneal cavity: The abdominal cavity formed by the peritoneum.
Per oral: By mouth.
Prostate: A gland surrounding the neck of the bladder in men.
Renal colic: Severe spasmodic pain in the loin bladder region.
Rigors: Feeling of cold with shivering accompanied by a rise in temperature.
Seminal glands: A gland producing a component of semen.
Septicaemia: Blood poisoning.
Serum: The clear portion of the blood.
Sinus: A fistula to a deep abscess.
Spall: A splinter or chip of rock.
Uraemia: Poisoning of the system by urinary matter.
Ureter: The duct by which urine passes from the kidney to the bladder.
<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urethra</td>
<td>The duct by which urine is discharged from the bladder.</td>
</tr>
<tr>
<td>Urolithiasis</td>
<td>Urinary stone disease.</td>
</tr>
<tr>
<td>Urosepsis</td>
<td>Poisoning of the urinary system.</td>
</tr>
<tr>
<td>Vesicle pain</td>
<td>Pain in the bladder.</td>
</tr>
</tbody>
</table>
The introduction of extracorporeal shock wave lithotripsy in the early 1980's possibly represents one of the most significant advances in the treatment of urinary and biliary stone disease, in the 3000 or so years of its known existence and treatment by man. Until relatively recently, patients often suffered appalling injury at the hands of the lithotomists who cut for the stone. In the review of stone disease, (chapter 2) an attempt has been made to describe the cause of the disease and plot the changing medical treatment through out the centuries to the present day. Despite the use of urethral dilators as early as Egyptian times and lithotrites (instruments passed through the urethra to crush or drill the stones) a truly non invasive technique had to wait for the advent of modern ultrasonics and the exploitation of piezoelectricity. This most notably occurred during world war I in the development of ASDIC, the allied ultrasonic echo ranging submarine detection system. Early experimental work on the use of ultrasound for the disintegration of urinary and biliary stones was carried out by workers such as Berlinicke and Schennetten1 (1951), Mulvaney2 (1953) and Coats3 (1956). Unfortunately, their results indicated that ultrasound induced stone disintegration only occurred at the sort of energy densities that also caused severe tissue damage. The breakthrough came in a technique first published in 1974 by Forbmann4 and co-workers, in which shock waves were used to fragment the stones. The acoustic shock waves were generated by an under water spark discharge between two electrodes at the first focus of a semi-ellipsoid reflector. This technique formed the basis from which Dornier GmbH produced the first commercial lithotripter5.

More recently a second and third generation of lithotripters have been introduced, some of which are based around the electromagnetic acoustic transducer or EMAT. The EMAT, which has been developed from the principle of the electromagnetic shock tube6 first described by Eisenmenger, operates in a similar way to that of a piston acting on a liquid. A high current capacitor discharge through the EMAT's flat spiral coil, results in the displacement of a metallic diaphragm which radiates a planar acoustic wave into the surrounding medium. During propagation, the wave becomes progressively more distorted due to the non linear effects of the medium, eventually taking on the saw tooth form of a shock wave. The addition of a focusing acoustic lens increases the pressure in the focal zone of the device to levels suitable for use in
lithotripsy. As a result of these increased pressures, particularly the negative pressure associated with the rarefaction phase of the focused shock wave, cavitation is often produced. This is seen in the acoustic field as both bubbles and cavitation transients due to the collapse of some of the cavities. The construction of an EMAT similar to that described by Reichenberger and Naser\textsuperscript{7} (1986) and a focusing lens is discussed along with the design of a piezoelectric hydrophone in chapter 3. Using the hydrophone, the acoustic field generated by the EMAT has been characterised. The results show that the focal region has a wider pressure distribution in the direction of propagation than in the transverse direction.

Studies in the area of shock wave interaction with urinary and biliary stones have identified several mechanisms such as spalling (Lubock\textsuperscript{8} 1989), cavitation (Sass et al\textsuperscript{9} 1991) and liquid jets (Crum\textsuperscript{10} 1988) that appear to collectively contribute to the disintegration of the various types of stone. Even now the role of any of these mechanisms in the eventual break-up of the stone is not fully understood. The problem is further confused because their individual contribution appears to be dependent on the type, (Dretler\textsuperscript{11} 1988) shape and location (Delius and Gambihler\textsuperscript{12} 1991) of the stone. In order to understand these processes we require a high resolution imaging technique, capable of not only imaging the stone and cavitation bubbles, but the shock wave and the cavitation shock transients that appear to have such an important role in the destruction of the stones.

High resolution optical interferometry is an ideal method for studying the path length and hence the density distribution of shock phenomena. However the study of spatially large events with methods such as Mach-Zehnder interferometry, require complicated arrangements of expensive large area optics. A less complex and cheaper alternative is to use Schlieren methods. The basic idea behind the Schlieren method and its use for observing optical disturbances was described by Toepler\textsuperscript{13} in 1867. Since then numerous workers such as Rayleigh\textsuperscript{14}, Scharin\textsuperscript{15}, Barnes and Bellinger\textsuperscript{16} have written on this subject. During this period the method has been applied to such phenomena as air flow analysis, combustion and plasma density measurements. The method yields both quantitative and qualitative information about an optical inhomogeneity. It is this and the ease with which large fields of view can be imaged that prompted its use for the study of not only the EMAT's acoustic field but the shock wave interaction with simulation stones.

Schlieren methods essentially rely on the deflection of a ray of light from its undeflected position by an optical inhomogeneity. The degree to which the ray is
deflected and thus the illumination changed in the image plane of the schlieren system, depends on the strength of the refractive index gradient introduced by the inhomogeneity. In the case of a shock wave, these refractive index changes are brought about by the local change in density associated with the rapid pressure fluctuations of the wave. Visualisation of the inhomogeneity depends on the particular schlieren method chosen, amongst the many variants are the shadowgraph method, the Toepler method and the scale method. The shadowgraph method allows the linear displacement of the ray in the image plane to be determined, whilst with the Toepler method, the angular deflection of the ray is obtained. In the scale method a transparent grid or scale is positioned in the object plane of the schlieren system. When the grid is viewed through the inhomogeneity the grid lines appear displaced. In each case Quantitative information about the acoustic field can be obtained by measuring the displacement or angular deflection of the rays making up a particular part of the schlieren image. A more detailed look at the theory and use of each of these methods can be found in chapter 4.

Prior to presenting the results for the schlieren observation of the EMAT acoustic field in chapter 5, the thesis describes the experimental apparatus. The last part of chapter 4 is primarily concerned with the overall equipment layout, specification and the control systems of the EMAT and schlieren set up. The schlieren and shadow images of the acoustic field presented in chapter 5 show the complex structure of not only the main shock wave, but also that associated with cavitation in the field. By comparing the transducer measurements with the quantitative schlieren images generated using a knife edge and opaque spot spatial filters, it has been possible to relate a specific component in the schlieren image to its transducer counter part. This has allowed the identification of the components within the images due to the negative or rarefaction phase of the shock wave and those that are due to regions of compression. The quantitative data has also allowed an assessment of the shock wave pressure to be made. One of the most interesting features of the images that will be presented is the non symmetrical appearance of some of the cavitation bubbles and shock transients. It is suggested that the lack of symmetry exhibited by the cavitation is an optical lensing effect introduced by the shock wave. To test this theory the scale method was used to model the distortion. The technique was also used to quantitatively analyse and partially reconstruct the EMAT shock wave. A high resolution study utilising sequences of individual schlieren images showing a shock wave during its interaction with simulation stones and other targets will also be presented. In the case of the simulation kidney stones, the images allow a clearer interpretation of the events producing the observed stone damage. The thesis concludes (chapter 6) with a look
over the research as a whole, commenting on the results and the suitability of schlieren methods for observing shock phenomena.

References

15 Schardin, H., Schlieren methods and their applications. R.A.E Translation 122.
CHAPTER 2

A HISTORY OF THE STONE

2.1 Introduction

Thomas Sydenham (1624-1689), 'The English Hipocrates' wrote with personal feeling when he described the symptoms of 'stone' thus, 'He suffers until at last he is worn out by the joint attack of age and disease and the miserable wretch is so happy as to die'. The story of the stone or calculus is one of the longest in medical history, it records the stoicism of the patients who submitted themselves to the agony of surgery in order to escape the tortures of the stone and the 'surgeons' endeavour in developing the techniques of cutting for the stone. The disease of urolithiasis has possibly plagued man since the dawn of time, the operation to remove the stone is speculated to be one of the first operations performed by man for the relief of a specific surgical condition. Circumcision and trephination of the skull are probably of comparable antiquity, but these were carried out for religious and superstitious reasons. It is known that cutting for the stone was practised in ancient India and Egypt from a very early date. In ancient Greece (450 - 350 BC) it is clear that the Hippocratic physicians recognised the symptoms of the stone and that lithotomy was practised by individuals who specialised in the operation.

The Roman writer Celsius in the 1st century AD described in his De Re Medicina an operative procedure in which, a cut was made through the perineum in the mid line down to the stone, after it had been manoeuvred into the neck of the bladder by rectal and abdominal palpation. Even the renowned Greek surgeon Galen, court physician under Marcus Aurelius in the second century AD offered no alternative and no alteration to the procedure. The decline of the Roman empire around the 4th century AD and the take up of Graeco-Roman medicine by the Arabians, left the procedure much as original and in fact the methodus Celsiana or apparatus minor as it was later known, was still the operation of choice until the 16th century. Throughout the medieval period lithotomy was performed by itinerant 'specialists', who although generally lacked any medical qualifications, served a useful purpose in the community. These specialists were often more skilled than the trained surgeons, some of the procedures they carried out were carefully guarded secrets, any improvements in the techniques often stayed within a family of lithotomists. However, because of the lack
of education and regulations governing the lithotomists, incompetence was common place and often cancelled the efforts of the few skilled lithotomists. Those that survived the procedure to remove the bladder stone would invariably suffer from impotence, incontinence and persistently draining sinuses. The oldest bladder stone recorded was that discovered in the grave of a 16 year old boy in a prehistoric cemetery at El Amrah upper Egypt. The pottery found in the grave suggested the boy lived in the predynastic period i.e. the middle or late middle prehistoric period, several generations before the rule of Menes, the first dynastic king c. 4800 BC. The calculus had a maximum diameter of 6.5 cm and consisted of a laminated phosphate crust with a uric acid nucleus. The incidents of stone disease in Egypt in recent times is notorious and was connected by many writers, Melton 1902 and Alphinus (De Medicina Egyptorum 1719) with Bilharziasis (Schistosomiasis), a small flatworm which settles in the veins of the urinary bladder or intestine, common by the banks of the Nile. When Napoleon's troops invaded the Nile delta in 1799, they were invaded in turn by the Schistosomiasis parasite. Napoleon, on seeing the symptoms of the disease, haematuria, (the presence of blood in the urine) was said to have called Egypt 'the land of menstruating men'. Professor Shattock speculates whether an epidemic of Bilharziasis and specifically, Bilharzial haematuria due to contact with water from the Nile through drinking or bathing, could be an explanation for the first of the biblical plagues (Exodus 7). He concedes that the calculus itself showed no sign of the presence of Bilharia ova and that a reference to the narrative itself shows the problems associated with such an idea.

The pages of history are littered with sufferers of the stone, of the famous, philosopher Bacon, scientist Newton, physicians Harvey and Boerhaave, the anatomist Scarpa, the writer Horace Walpole, Peter the Great, Louis XIV, George IV, Oliver Cromwell, Napoleon III are but a few. The best known stone is that of Samuel Pepys, 'I remember not my life without the pain of the stone in the kidneys (even to the making of bloody water upon any extraordinary motion) until I was 20 years of age'. Whilst a student at Trinity Hall Cambridge (1653) Pepys suffered from an attack of renal colic, after which he was subject to violent attacks of vestical pain. The bad winter of 1658 brought things to a head and he was cut for the stone on the 26th March by Thomas Hollier of St Thomas' hospital. He was fortified for the operation by a draught containing liquorice, marshmallow, cinnamon, milk, rose water and whites of eggs. The bladder stone was tennis ball sized, weighing about 57 grams (2 oz) and consisted mainly of urates. The relief from the tortures of the stone profoundly changed his life, for years to come he and his friends would feast on the anniversary of his delivery, he called the celebration 'my solemn feast for the cutting of the stone'. Napoleon
Bonaparte, who was another sufferer, had frequent attacks of dysuria. At the battle of Borodino in the Russian campaign of 1812 he had to dismount from his horse frequently to pass water which contained 1/3 sediment. In exile on St Helena he was seen at times with his head against a wall or tree passing urine in small, painful dribbles. He wrote, 'This is my weak spot, it is by this that I shall die.' On his death, aged 51 on the 5th May 1821, a postmortem was carried out by Dr Dominique Antommarchi. He noted among other things that not only did the bladder contain quantities of gravel mixed with small calculi but that the bladder itself was in a diseased state. There is, however, evidence to suggest that Napoleon died not as a result of the diseases associated with the stone, but of symptoms attributed to arsenic poisoning. The suspected source was the green copper arsenite pigment (Scheele's green and paris or emerald green) present in the wallpaper at longwood house. One of the most remarkable cases of lithotomy was that of Jan de Doot a Dutch blacksmith and lithotomist who, in 1651, surgically removed a 113 gram (4 oz) stone from his own bladder using the supra-pubic extraperitoneal approach, which will be discussed later. This stone was certainly not the largest ever reported in England, this record belongs to a stone that was removed in 1809 which weighed 1.36 kg (2¾ lb). An equally large stone weighing 1.13 kg (2½ lb) was removed in a London hospital in 1975.

Kidney stones (Nephrolithiasis) were almost certainly as common as bladder stones in antiquity, these too had their famous victims; Pope Innocent XI (1611-1689), who struggled continuously against the absolutism of Louis XIV in church affairs, suffered from the stone, his kidneys containing enormous calculi. Although the clinical picture was certainly recognised, the impossibility of the operation and the certain patient mortality prior to the advent of antiseptic surgery, prevented any advancement similar to that achieved with the procedures for the bladder stone. The breakthrough came in 1867 when Joseph Lister published a paper on antiseptic techniques in surgery. Shortly afterwards the first planned nephrectomy was performed by Gustav Simon of Heidelberg in 1869, using anaesthetic and antiseptics. The story of the gallstone is similar to that for the kidney stone, with the first cholecystolithotomy (the removal of a stone through an incision in the gallbladder) being performed in 1867.

2.2 Stone disease

The growth of stones or solid masses generally occurs in small cuplike cavities and similar structures within the body. The common sites for stones, a variety of which are
shown in figure 1, are the urinary tract (kidneys, bladder), the biliary system (gallbladder, bile ducts) and sometimes the salivary ducts. In each case any one of a number of factors may be associated with the formation of a stone. The secretions produced by various glands within the body, can by a gradual process of sedimentation form solid masses. Examples of these are cholesterol gallstones formed in the gallbladder, which are associated with cholesterol supersaturation of bile and uratic stones found in the kidneys, or bladder of a person suffering from acidic urine. It is also responsible for the plugs of hardened wax in the ear and the cheese like masses which accumulate in the tonsils giving rise to bad breath. The deposition of lime salts in inactive i.e. damaged or degenerated tissue, is another factor associated with stone formation. Normal tissue produces carbon dioxide which prevents the deposition of lime salt from fluids circulating past or through them. Inactive tissue such as that seen in the healed up areas of the lungs after tuberculosis, or degenerated blood vessels, tumours, scars, the bodies of dead parasites and other foreign bodies, do not produce sufficient carbon dioxide to prevent deposition and in time may have masses of a considerable size formed on, or around them. A similar situation is seen with gout in which sharp uratic crystal are deposited on the surface of joints and in other tissues resulting in the eventual formation of chalk stones. Bacterial action can also result in the deposition of lime and phosphates, such as tartar on teeth, it is also responsible for some types of bladder stones and calculi found in the salivary ducts in the cheek or under the tongue.

Figure 1 Urinary stones.
We shall now look at the particular conditions associated with the formation of the various types of stones and comment on the disease symptoms and method of treatment. To do this it is first necessary to understand the structure and function of the human urinary and biliary systems.

2.3 The urinary system

The metabolism of proteins and other nutrients in the body produce carbon dioxide, water, heat and toxic nitrogenous wastes such as ammonia and urea. It is essential if poisoning of the system or uraemia is to be prevented, that all the toxic material and excess essential ions such as sodium, chloride, phosphate and hydrogen are removed from the body. The primary function of the urinary system (figure 2) is to maintain the body in a stable equilibrium or homoestasis, by controlling the quantity of these substances circulating in the body. The process begins in the kidneys, which are a pair of glands positioned either side of the back bone. The kidneys are surrounded by a quantity of fat and loose connective tissue, in which are embedded the large vessels which supply them with blood. They are protected by the thick muscles of the back, abdomen and their upper halves by the 11th and 12th ribs. Each kidney is approximately 10 cm long, 6.5 cm wide and weighs around 140 grams, but depends to a large extent on the body weight and the habits of the individual. The left kidney is slightly longer and narrower than the right which is a little lower in the body than the left due to the liver. The spleen, pancreas and stomach lie in front of the left kidney and the duodenum and several loops of the intestine in front of the right.

Figure 2 The male urinary system.
Chapter 2: A history of the stone

The urine formed by the kidneys enters the renal pelvis, a funnel shaped structure that narrows to a fine tube known as the ureter. These pass down the back of the abdomen a distance of 25 cm and enter the base of the bladder through the ureteral orifice. The bladder lies deep in the pelvis behind the pubic bones, in front of the bowl and is suspended by numerous ligaments. The wall of the bladder is made of muscular and fibrous tissue which run in various directions to give the wall strength, its innermost side is lined with a mucus membrane. The two are loosely coupled to allow expansion, when fully expanded the bladder rises up into the abdomen and has a capacity of around 570 ml. A third opening, through the prostatic urethra leads to the exterior, via the urethral orifice, a distance of 4 cm in women and 20 cm in men. The exit of the bladder is kept closed by a muscular ring which is relaxed every time water is passed.

2.31 The kidney

The chief function of the kidney is to regulate the composition and volume of blood by separating from it, fluid and certain solids. The blood enters the renal pelvis of the kidney through very large renal arteries, themselves branches of the principle artery of the body, the aorta. The renal arteries carry approximately a quarter of the blood circulating in the body, resulting in a total flow through the kidney of around 70 litres an hour. The working unit of each kidney is the nephron figure 3, of which there are about a million. Each nephron begins with a small branch from the renal artery, which divides to form bundles of convoluted tubules (medullary rays), each of which end up in a small rounded body, the malpighian corpusle or glomerulus. A long winding tube, the renal tubule, is headed by a goblet shaped extension (bowmains capsule), which completely surrounds the glomerulus. The capsule is permeable and serves as a collecting chamber for the fluid filtered by the glomerulus. After circulating through the glomerulus the concentrated blood emerges by a small vein which splits up again into capillaries on the walls of the renal tubule, where certain solid contents are removed. The blood is then collected and leaves the kidney through the renal vein.

The glomerular filtrate, which can be up to 10 % of the blood volume passing through the glomeruli (120 cc per min or 160 litres a day) consists of everything in the blood except cells and proteins. Its worth noting that, at this rate the process would remove all the water in the body in about 5 hours along with 56 grams of sugar and 227 grams of salt. However in a healthy individual the tubule reabsorbes and returns to the blood 99 %, or, all but 1.5 lt of the 160 lt of filtrate. The rate of tubule reabsorption and thus the quantity of urine is controlled by a sensory organ in the hypothalamus, in the floor...
of the brain. In a healthy man this would be around 1.5 litres a day. In the event of a reduction in the dilution of the blood, due to excessive sweating, an anti-diuretic hormone (ADH) is released from the pituitary gland. This promotes tubule reabsorption and reduces the volume of urine therefore conserving water. However the quantity of urine cannot be safely reduced to less than 600cc per day, since this is the minimum amount that will carry away the waste products. Individuals who suffer from diabetes insipidus and lack ADH may pass as much as 20 litres of urine per day since they can only reabsorb around 90% of the glomular filtrate.

Another important function of the kidney is to regulate the acidity of the body fluids, which in a healthy individual is very slightly alkaline (pH 7.4). The bi-products of metabolism tend to be acidic, but is closely related to the diet. A high protein diet increases acidity, whereas a diet composed largely of vegetables would increase the alkalinity. To prevent a harmful build up of acid waste products in the blood (acidosis) the kidney acts in several ways. One of those is to produce ammonia which neutralises the acid to form ammonium salts. Another is the production of acid phosphate from alkaline phosphate taken from the blood. The kidney also secretes hormones such as renin which can raise the blood pressure and esythopoientin that stimulates the formation of red blood cells.

![Figure 3 Section of the kidney and the nephron right.](image)

The constituent of the filtrate can be grouped by the rate at which they are reabsorbed by the tubules.
1) Substances actively reabsorbed, such as amino-acids, glucose, sodium, potassium, calcium, magnesium and chlorine.

2) Substances that diffuse through the tubule when the concentration in the filtrate exceeds that in the plasma, i.e. urea, uric acid and phosphate.

3) Substances not returned to the blood from the tubular fluid e.g. creatine, a by product of protein metabolism.

The remaining fluid or urine in the tubules is carried away to collecting ducts and on to the renal pelvis, where it leaves the kidney.

2.32 The renal calculus or kidney stone

The kidney can tolerate a great deal of damage, even in cases of advanced kidney disease the symptoms are often mild and cause no more than a sense of not being well. The analysis of the urine (urinalysis) for the presence of substances not normally expected to appear, or even the presence of normal constituents in abnormal amounts in the urine will generally give an extremely accurate indication of the disorder. For example the presence of sugar in the urine or glycosuria can be an indication of the liberation of glucose from the liver due to emotional stress, or the inability of the pancreas to produce sufficient insulin, as with diabetes. Other examples are the presence of leucocytes and erythrocytes, i.e. the components of pus (pyuria) and the appearance of red bloodcells (haematuria) in the urine. Among other things these can both indicate acute inflammation of the urinary organs due to a kidney stone. The presence of a stone somewhere in the urinary tract may also be given away by the presence of its crystalline deposit in the urine. Of the disorders associated with the urinary tract, kidney stones are the major cause of morbidity. They cause considerable suffering and have a substantial economic impact, in 1986 more than $2 billion was spent on the treatment of kidney stones\textsuperscript{10}. Stones are more common in men, and studies suggest that the incidence is on the increase, not only in the United States\textsuperscript{11,12} but Sweden\textsuperscript{13} and Japan\textsuperscript{14}. It is estimated that around 2% of the population in the United Kingdom may have calculi and that the incidence in other parts of the world, particularly the middle east, is much higher.

In most cases stones are formed by the deposit in the urinary passages of solid substances naturally present in the urine. There deposition may be associated with
Chapter 2: A history of the stone

their presence in excessive amounts, or to the failure of some process that keeps them in solution, or simply due to bacteria. The most common constituents of stones are ammonium phosphate, calcium phosphate, calcium carbonate, calcium oxalate, uric acid and urates. Some of the common stones found in the urinary system along with their composition and structure are listed in table 1.

<table>
<thead>
<tr>
<th>Type</th>
<th>Structure and Linear Density (g/cm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calcium Oxalate Dihydrate</td>
<td>Delicate crystal structure (0.715 g/cm²)</td>
</tr>
<tr>
<td>Calcium Oxalate Monohydrate</td>
<td>Uniformly dense (0.815 g/cm²)</td>
</tr>
<tr>
<td>Calcium Phosphate Dihydrate (Brushite)</td>
<td>Dense laminated structure (0.837 g/cm²)</td>
</tr>
<tr>
<td>Calcium Phosphate (Aptite)</td>
<td></td>
</tr>
<tr>
<td>Cystine</td>
<td>Homogenous crystal structure (0.402 g/cm²)</td>
</tr>
<tr>
<td>Uric Acid</td>
<td>Homogenous structure with concentric and radial laminations (0.226 g/cm²)</td>
</tr>
<tr>
<td>Magnesium Ammonium Phosphate (Stuvite/Aptite)</td>
<td>Layered accumulation of material (0.449 g/cm²)</td>
</tr>
</tbody>
</table>

Table 1 Common urinary calculi.

Many factors are associated with, or contribute to, the formation of stones and it is some of these that will now be described. One factor suspected of being associated with stone formation is an abnormally high level of calcium in the urine (hypercalciuria), simply because of the number of stones that contain calcium. Changes in the acidity or alkalinity of the urine due to diet, as mentioned earlier is also associated with stone formation. An excessively acid reaction may be accompanied by the formation of uric acid stones, whilst phosphate stones form in alkaline urine. Calcium oxalate stones can form in individuals who don't synthesize the protein glycoprotein crystal growth inhibitor (GCI) in their urine, which prevents the growth of calcium oxalate crystals. Other factors associated with the formation of stones in the kidney are, prolonged recumbancy; urinary stasis, resulting from congenital or acquired obstructive lesions in the urinary system; infection of the kidney; excessive sweating, resulting in persistently concentrated urine; the lack of vitamin A in the diet, resulting in the increased incidence of urinary infection and gout, due to the presence of excess uric acid. Additionally, the overaction of the parathyroid gland leading to hypercalciuria and leukemia, resulting in uncontrolled infection due to the lack of mature or normal white blood cells, can contribute to the formation of stones. A grossly excessive intake of vitamin D can lead to the overabsorption of calcium and therefore also produce stones.

Apart from the obvious association of kidney stones with certain diseases, it is apparent that diet also has a major effect in the development of stone disease. A recent American study\textsuperscript{15} of 45,619 men has suggested that a high dietary intake of calcium,
which is strongly suspected of increasing the risk of kidney stones, actually has a beneficial effect and decreases the risk. This is all the more interesting, since patients with kidney stones are routinely put on low calcium diets\textsuperscript{16}. Patients with hypercalciuria put on this type of diet, have a 10% higher probability of stone formation\textsuperscript{17}, as well as the possibility of a negative calcium balance and bone loss\textsuperscript{18}. The study indicates that men who drank 240 ml of skimmed milk per day had a 40-50% better chance of not developing stone disease than those who drank less than 120 ml per month. Similar trends were also seen with cottage cheese, yogourt and non dairy sources of calcium such as oranges and broccoli. One of the possible explanations for the apparent protective effect of calcium, which may reduce the possibility of the formation of calcium oxalate stones, the most common, is the increased binding of calcium with oxalate in the gastrointestinal tract. Calcium oxalate saturation of urine increases rapidly with small increases in oxalate concentration\textsuperscript{19} and therefore urinary oxalate may be more important than urinary calcium for stone formation. The reduction in dietary calcium may therefore lead to increased oxalate excretion and thus the formation of oxalate stones. Whilst the study showed that the dietary intake of calcium and potassium\textsuperscript{20} appeared to have a beneficial effect in stone disease, this was certainly not the case for animal proteins. These were directly associated with the risk of stone disease due to the excretion of uric acid, as described earlier. The beneficial effect of increased fluid intake resulting in the dilution of urine is well known.

Surgical techniques are by far the most prevalent form of treatment in the world for stones in the upper urinary tract. Although these are now being rapidly replaced by minimally invasive techniques such as percutaneous endoscopic nephrolithotomy, extracorporeal shockwave lithotripsy and drug chemolysis. The move to minimally invasive techniques has greatly reduced patients morbidity, mortality and length of hospital stay. Prior to 1980 patients with kidney stones underwent 2 hours of open surgery and were left with a 25 cm long scar in the loin region. Open surgery is very traumatic and major complications are not unusual as is the loss of some renal function, because of this convalescence normally takes between 6-12 weeks. In 1979 percutaneous nephrolithotomy was introduced, this involves making a minor incision in the kidney through which an endoscope was passed. Small stones were simply removed, later developments allowed larger stones to be treated, by first disintegrating them with a ultrasonic or electrohydraulic probe. This treatment has dramatically reduced patient morbidity and mortality, as a result convalescence is possible within a week. Few complications\textsuperscript{21} are experienced other than post operative septicaemia, which can be treated with antibiotics. About the same time extracorporeal shockwave
Chapter 2: A history of the stone

lithotripsy (ESWL) was also introduced\textsuperscript{22}. This uses focused shockwaves to fragment the renal stones into small enough pieces such that they are passed naturally in the urine. Modern versions of the lithotripter are virtually painless, have virtually no major side effects and are able to treat 85\% of all simple renal stones\textsuperscript{23}. Of the remaining 15\% that are not suitable due to size or type, the majority can be treated by percutaneous nephrolithotomy. In 1\% of cases open surgery remains the only option.

The suitability of a stone for fragmentation with shockwaves will obviously be an important consideration when deciding on the treatment method. Diagnosis using radiography or an alternative technique can often yield information about the stones composition from its radiographic appearance and therefore influence the choice of therapy. For example a large calculus with small spaces suggests predominantly calcium oxalate dihydrate, whereas the smooth and dense radiographic appearance of a similar volume stone suggest calcium oxalate monohydrate. An in-depth look at lithotripter technology will be the subject of a later discussion. Other treatments include drug chemolysis and involves changing the pH of the urine. In the case of uric acid stones\textsuperscript{24,25,26,27}, the excessively acid urine is made alkaline by giving large doses of alkali, eg. 2g of sodium bicarbonate four times a day. In the case of phosphate stones the excessively alkaline urine is made acid\textsuperscript{28} by taking 1g of ammonium chloride four times a day. In cases of stones associated with hypercalciuria, interesting results have been obtained with bendrofluzide, one of a group of diuretics which act by inhibiting reabsorption in the renal tubules, also cellulose flouride and unprocessed bran. A summary of some of the features associated with the treatment alternatives is given in table 2.

| Criteria                  | Open surgery | Percutaneous lithotom | ESWL       | Drug/chemolysis |
|---------------------------|--------------|-----------------------|------------|-----------------
| Stone type                | any          | any                   | most       | Uric acid      |
| No of stones              | any          | any                   | ≤ 3        | Phosphate      |
| Stone size                | any          | any                   | ≤ 25 mm    | any            |
| Treatment duration        | 2-3 hours    | 1 hour                | 30-45 mins | ≤ 10 mm        |
| Hospitalisation           | 12-14 days   | 2 days                | out patient| months-years   |
| Convalescence             | 12 weeks     | 2-10 days             | immediate  | ---            |
| Anaesthetic               | general      | general               | none       | none           |
| Painful                   | yes          | slight                | sometimes  | sometimes      |
| Complications             | many         | septicaemia           | obstructions| ?              |
| Prob stone free           | ~ 100%       | ~ 100%                | ~ 90%      | ?              |

Table 2 Urinary stone treatment alternatives.

The passage of a calculus from the kidney through the ureter into the bladder can bring with it very severe problems. Sometimes a calculus can become lodged in the ureter, where it is described as an impacted ureritic calculus. Apart from the problems of
restricting the flow of urine from the kidney and the possibility of acidosis, excruciating pain known as renal colic may be felt. These pains are of an agonising nature and generally appear suddenly, shooting down from the kidney region into the groin and external genital organs.

2.33 The vesical calculus or bladder stone

Stones in the urinary bladder may have either formed there or, as described previously, migrated from the kidney. They are similar in composition to renal calculi and similar factors operate in their production. Calculi that remain in the bladder will normally increase in size due to the build up of further deposits from the urine. The most common type of stones found in the bladder are those composed of phosphates. These occur in long standing cases of inflammation of the bladder (cystitis). The inflammation may be associated with urinary obstruction due to an enlarged prostate in men, or the prolapse of the womb in women. Narrowing of the urethra and bacterial infection may also cause inflammation. The bacterial organisms most frequently found in the urine is *Escherichia coli* (E coli) which normally lives in the bowel. Infection and inflammation of the bladder is much more common in women because the urethra, vagina and anus are very close together and because the urethra is very short. *Mycobacterium tuberculosis*, which produces a chronic ulcer on the bladder wall and the schistosome, which has been mentioned earlier in this chapter and very common in Egypt and South Africa are also associated with cystitis. The other types of stone commonly seen are those composed of uric acid and oxalates. Among the symptoms, pain in the small of the back, haematuria, bad smell, rigors and the presence of a whitish sediment are common. In chronic cases the frequent and painful desire to pass small quantities of urine is experienced.

The increasing frequency of kidney stones in the western world is matched only by the rapid and intriguing fall in cases of the bladder stone. Throughout history there have been a large number of documented cases of the bladder stone. So common were they in previous centuries, that it was often a common cause of crying in infants. Today such cases in the developed world have virtually disappeared, mainly due to improvements in diet and hygiene. A survey by Anderson on the incidence of urolithiasis in Norway brought forward evidence to show that endemic bladder stones appear to be related to diets which are deficient in animal products and rich in vegetable proteins. The majority of cases now occur in older age groups and in
underdeveloped areas of the world where vegetable protein makes up a substantial part of the diet, eg India, Thailand and China.

Once a bladder stone has been diagnosed, treatment often becomes necessary. This is generally achieved surgically (lithotomy), although crushing the stone (lithotritry) with an instrument known as a lithotrite has also been used. In the last 20 years percutaneous endoscopic lithotomy has gradually replaced open surgery, which is now only necessary for extremely large stones. Apart from the clinical differences, the procedure is virtually the same as that used for the treatment of kidney stones. The use of ultrasound (lithotripsy) for the destruction of kidney stones is well known, however it is not known whether this technology is suitable or regularly used in the treatment of bladder stones. Early attempts at chemical dissolution of stones, using lithotryptics, by the ancient Indians and the Greeks proved unsuccessful. In the 18th century, stone dissolving chemicals were once again actively sought, because of the mortality and side effects associated with lithotomy, however they were rarely effectual. One such stone dissolving remedy was prescribed by Mrs Joanna Stephens in 1739. It was thought so effective by doctors of the day that the government, through a special act of parliament purchased it for £5000 for the nation. Not surprisingly it didn't work, it was later found to contain calcined egg shells, old tobacco pipes, soap and aromatic bitters, i.e. lime, phosphates and alkalis. The operative procedures used for the removal of bladder stones has had a long and chequered past it is these and their patrons that shall be discussed next.

2.34 Early operative procedures

2.341 Perineal lithotomy

Perineal lithotomy is the collective name given to procedures in which, bladder stones were removed from the bladder via an incision in the region between the anus and the scrotum or vulva, the perineum. The operative procedure described by Celsus and later known as the apparatus minor used such an approach. The operation required very few instruments, generally all that was required was a knife and a hook or a pair of forceps to extract the stone, hence the name. The procedure had been described by many surgeons throughout the 1500 years it was in use. As well as Celsus, the Hindu surgeon Susruta of Benares c. 6BC- 6AD, commented that the surgeon was first to ensure that his fingernails were closely cut. The German surgeon Lorenz Heister (1683-1758) also gave a detailed description of the procedure. He noted, as Celsus,
that the operation was only suitable for boys under the age of 14, since the procedure involved opening the base of the bladder just above the prostate and only in boys below this age, are the pre pubertal prostate and seminal glands small enough to allow the operation. Several days before the operation the patient was required to fast and immediately prior given an enema, in which gas or liquid was injected into the rectum, to expel its contents. The child was then held in the lithotomy position, in the lap of a strong and intelligent person! and the operation begun. Initially the stone was maneuvered into position by inserting the well lubricated fore finger of the left hand into the anus, whilst pushing down onto the bladder with the right. The stone was then held between the fore finger in the rectum and the perineum, producing a bulge as figure 4. A slightly curved cut was then made down onto the bladder, which was cut just above the prostate. The stone was then pushed out using the fore finger and if necessary helped using the hook or forceps, the wound was then dressed with wool and warm oil. The operation was rarely satisfactory and amongst other complications the wound in the bladder floor often did not heal correctly leaving a chronic fistula.

Figure 4 The apparatus minor.

The first important change to the procedure occurred in 1535, when Marianus Sanctus Barolitanus (1490-1550), published a technique developed by his teacher, the Italian surgeon Fraciscus de Romanis of Cremona. The modification is said to have been inspired by the relative ease with which it was possible to remove a stone from the female bladder by first dilating the urethra. It might be worth reminding the reader here that the length of the urethra in women is approximately 4 cm as opposed to 20 cm in men. The Marian operation or apparatus major as it was known, because of the need for additional instruments, required a certain degree of precision. The operation required a grooved staff to be passed into the bladder through the urethra, which was then used to guide the operator to the bladder neck. An incision was then made in line with the groove through the urethra and a dilator introduced, the prostate and bladder neck were then forcibly torn and the stone removed using forceps. The operation was
of course very traumatic, haemorrhaging and damage to the rectum were frequent fatal complications and incontinence and impotence following the operation were not uncommon. It would be 300 years before the discovery of anaesthetic and aseptic precautions, hence all that a patient could reasonably expect for the pain was perhaps opium or alcohol and possibly the presence of a churchman. Throughout the operation the patient had to be kept still, to aid the surgeon therefore the patient would be bound and held by three or four strong men in the lithotomy position, i.e a similar position to that shown in figure 4. The next stage in the development of lithotomy came with the introduction of the lateral approach by Hugenot Pierre Franco (1500-1561) one of the stars of renaissance surgery. Although he received no formal medical education, he rejected the brutal method of tearing the prostate and bladder neck, preferring instead to make an incision through the same. He was also responsible for the first recorded suprapubic lithotomy in which the bladder was approached via a lower mid line abdominal incision. In addition he was also credited for his pioneering work in the area of hernia, ophthalmic and facial plastic surgery. Further advancement in the lateral approach came with an unqualified itinerant french lithotomist call Jacques Beaulieu31 (1651-1714). Who in 1690 changed his name to Frère Jacques and adopted the habit of a monk, he is also unique in that his name has been continued in the form of the nursery rhyme, Frère Jacques, Frère Jacques, Dormez vous ? Dormez vous ? .............. Ding dong ding.

In 1697 he applied to cut for the stone at the hotel Dieu, paris. Despite a satisfactory operation to remove the stone from a corpse, the board of surgeons refused his licence, possibly due to the fact that no pre operative bleeding or purging was carried out and because of his refusal to reduce bleeding with an astringent, prefering instead to leave it to God. He continued to practice with varing degrees of success, this aside, he was still a very able operator and an extremly honest and charitable man. His early failures were due to his general ignorance of anatomy, which he remedied during a period of collaboration with Fagon, surgeon to king Louis XIV. He is said to have operated on nearly 5000 people with bladder stones and a further 2000 for hernias.

The last major modification to the lateral approach before its ultimate demise and succession by the suprapubic approach was made by William Cheselden32,33 (1688-1752). Cheselden an expert anatomist and possibly the greatest lithotomist in the world, operated swiftly with great precision. The modification in 1725 involved opening the bladder by way of the prostatic urethra as shown in figure 5. The combination of speed and accuracy resulted in a mortality rate of only 9% in 213 patients. In fact in his early operations mortality was less than 6%, Cheselden
explained that initially only fit patients expect to survive were given the operation. Later when there was greater demand for the operation even the most aged and miserable cases expected to be saved by it.

Figure 5 The lateral approach.

It is worth noting that the lowest mortality rate contemporary lithotomists could achieve at that time was around 40%. Throughout the time perineal lithotomy was being used to treat bladder stones, the search for a noninvasive method to avoid the operation continued. There was little success with Lithotryptics, but the removal of stones or fragments of stones through the urethra (lithotrity) had been practiced since antiquity.

2.342 Trans urethral lithotrity

In the late part of the 18th century lithotrity was reintroduced and used alongside Cheseldens modified lateral lithotomy. From early times the Egyptians were known to have dilated the urethra using a wooden cannula. Other patients were known to have experimented on themselves, in one such case a man introduced a long nail into his bladder. On locating the stone with the end of the nail, the other end was struck hard with a blacksmith's hammer. General Martin of Lucknow in 1783 was said to have gradually, over a nine month period, disintegrated his bladder stone using a metal
sound with a roughened edge. The man credited with performing the first successful lithotrity was the Frenchman Jean Chiviale (1792-1867) in 1824. He performed the operation at the Necker hospital in Paris, using instruments of his own design. Chiviale was later to instruct Henry Thompson, a surgeon at University College Hospital in the art of lithotrity. Thompson using lithotrity, was to later treat two royal bladder stones, those of Leopold I of Belgium in 1863 and Napoleon III in 1873, who, unfortunately succumbed during the course of treatment and died, due to his general state of ill health.

Early lithotrites were in some cases very complex and ingenious instruments, they generally consisted of a mechanism to ensnare the stone, such as a wire loop or a multi-fingered arrangement, through which was passed a gimlet or drill, figure 6a. At each of the several sittings that were required, figure 7, the stone would be hollowed out or perforated until it was eventually broken up into small enough pieces to pass through the urethra in the urine per natura. Other instruments, such as Jacobson's articulated stone crusher (1829) figure 6b, crushed the stone between two blades brought together by a screw action. In Heurteloup's percussion lithotrite, the calculus was seized between the teeth of the instrument which was fixed to the operating table. The stone was then crushed between the teeth by repeatedly hitting the end of the instrument with the hammer.

Figure 6 a) Chiviale's trilabe 1824, b) Jacobson's articulated stone crusher 1829.

Lithotrity, although a safer procedure than perineal lithotomy had several drawbacks. The presence of stone fragments in the bladder after crushing often caused problems as did the reliance on the passage of the fragments naturally in the urine. The presence of stone fragments often led to inflammation of the bladder and urethral obstruction, causing frequent urination, pain and haematuria. The first steps to improve the procedure came with an idea of Sir Philip Crompton (1777-1858), in which he devised a method of removing stone fragments from the bladder using a steel catheter and a
heavy metal or glass suction bottle. Rapid lithotripsy or litholapaxy allowed the stone to be crushed and its fragments removed in one operation, instead of the multiple sittings required for lithotrity. The operation wasn't truly established until 1878, when Jacob Bigelow (1818-1890) professor of surgery at Harvard University, published 'Lithotrity by a single operation'. Despite this advance and the obvious advantage it had over operative techniques, it never became popular, Surgeons preferring the high operation for the stone or suprapubic lithotomy.

Figure 7 Patient undergoing trans urethral lithotrity.

2.343 Suprapubic lithotomy

The first recorded attempt at the use of the suprapubic approach was carried out by Huguenot Pierre Franco on a child of about 3 years old, sometime before his death in 1561. This was tried only after the perineal approach had failed due to difficulties in getting the large stone into the neck of the bladder. The operation was fortunately a success, but he advised other surgeons not to copy, due to the risk of urine entering the peritoneal cavity from the incision in the bladder and the possibility that the intestines would prolapse through the abdominal wound. By 1717 the operation was once again being considered, John Douglas realised that the bladder could be opened easily when fully distended and that the operation also avoided all of the common complications seen with perineal lithotomy. William Cheselden was also a one time
proponent of the approach, which he described in a book published in 1723. However a year later he returned to his work on improving the lateral perineal approach, when it was obvious this operation had its own serious complications. It wasn't until the introduction of general anaesthesia (1846) and asepsis (1867) that the suprapubic approach was used routinely and safely, and developed into the procedure used today.

2.4 The biliary system

The biliary system (figure 8) is composed of the gallbladder and the hepatic and common bile ducts and are among a group of organs or accessory structures that aid the physical and chemical breakdown of food in the gastrointestinal tract. Grouped together the organs of the gastrointestinal tract, such as the mouth, pharynx, esophagus, stomach, small and large intestine and the accessory structures, such as the tongue, salivary glands, liver, gallbladder and pancreas, form the human digestive system.

The main function of the biliary system is to store and concentrate in the gallbladder, bile secreted by the hepatic cells of the liver and when required discharge it via the common bile duct into the duodenum or small intestine. The liver, the largest organ of the body, weighing about 1.4 kg in the average adult, is located under the diaphragm occupying most of the right hypochondrium (the part of the abdomen covered by the cartilage of the lower ribs). It is divided into two principle lobes, the left and right. From each of these lobes bile capillaries carry bile synthesised by the liver cells to ducts which merge to form the left and right hepatic ducts, these eventually unite and leave the liver as the common hepatic duct. This is then joined a little further down by the cystic duct from the gallbladder and becomes the common bile duct. The gallbladder is a pear shaped sac, about 7 to 10 cm long resting on the underside of the right lobe of the liver. The common bile duct and pancreatic duct then join to form a common duct known as the hepatopancreatic ampulla or ampulla of vater, which enters the duodenum on an elevation known as the duodenal papilla, 10 cm from the exit of the stomach. The liver performs many of the vital functions necessary to regulate the physical and chemical composition of the bodies internal environment. Among these are the regulation of sugar, lipids and amino acids, the formation of cholesterol and red blood cells and the elimination of sex hormones and haemoglobin, as well as the storage of up to 1.5 l of blood and the production of heat and bile.
Bile is partially an excretory product containing waste removed by the liver and partly a digestive secretion. Each day the liver produces between 800-1000 ml of the thick, bitter, brownish or olive green, alkaline (pH 7.6 - 8.6) liquid. Its constituents are typically water, mucus, pigments (bilirubin and biliverdin), salts of three complex acids (cholic, chenodeoxycholic and deoxycholic acid), cholesterol, lecithin and some mineral salts. The digestive components are the bile salts, such as sodium taurocholate and glycocholate, emulsify fat globules by reducing their surface tension. The break-up of fat globules in this way produces a suspension of fat droplets of around 1μm in diameter, this increase in surface area improves the digestive action of the enzyme lipase, a constituent of gastric juice. Bile is also rich in sodium bicarbonate, which neutralises stomach acid and produces a favourable alkaline environment for the various enzymes in the small intestine. The principle bile pigment, bilirubin, is produced when haemoglobin in used red blood cells is broken down by the liver. This brown pigment is further broken down in the small intestine to another pigment that is present in faeces and which gives them their characteristic colour. Bile also contains cholesterol, this fat derivative is an important constituent of cell membranes, particularly nerve cells, any excess in the body is generally excreted by the liver. Cholesterol is made soluble in bile by the bile salts and lecithin.

The bile secretion rate is relatively constant, however there are certain periods during the day, such as at meal times, particularly after a fatty meal, when additional bile is
required. This additional requirement is usually supplied by the gallbladder. Usually when the intestine is empty a valve or the sphincter of the hepatopancreatic ampulla is closed, preventing bile entering the intestine. Since bile is being secreted constantly it begins to back up, eventually over flowing into the gallbladder for storage until it is required. Whilst in the gallbladder the absorption of water and many ions results in anything up to a ten fold increase in its concentration. Shortly after taking a meal, acidic semisolid and partly digested food (chyme) enters the duodenum. If particularly high concentrations of fat or partly digested proteins are present, the intestinal mucosa secrete a hormone known as cholecystokinin. This stimulates the muscular contractions of the gallbladder walls and also relaxes the sphincter of the hepatopancreatic ampulla, allowing concentrated bile into the intestine to assist with the processing of fat. As the chyme continues its passage through the small intestine (typically at a rate of approximately 1cm / minute), 90% or the major part of digestion and absorption will take place. Final absorption takes place in the large intestine, where the faeces is also formed and eventually expelled through the anus.

2.41 The biliary calculus or gallstone

The two most common disorders affecting the biliary system are cholecystitis or the inflammation of the gallbladder and gallstones. Cholecystitis is often due to infection by microbes and very infrequently due to a complication of typhoid or paratyphoid fever. In acute cases, inflammation can develop due to the blockage of the cystic duct by an impacted gallstone. In these cases fever, increased pulse rate, pain in the right upper quarter of the abdomen and a characteristic tenderness of the right tip of the ninth rib is apparent. In chronic cases of inflammation, indigestion, flatulence, intolerance to fatty foods, pain in the right shoulder blade and periodic attacks of vomiting are common symptoms. The elderly are particularly prone to additional complications such as the perforation of the gallbladder wall resulting in the formation of a fistula with the small intestine. Gallstones are also almost exclusively associated with carcinoma of the gallbladder, fortunately this only occurs in 0.4% of cases.

Gallstones are typically formed as a result of precipitation from the bile of one or more of its main constituents. These insoluble deposits form three main types of stone, cholesterol, bile pigment and mixed stones consisting of cholesterol, bile pigment and calcium. As with urinary stones, gallstones will often form around foreign bodies such as suture material or the corpses of dead parasites present in the bile ducts or gallbladder. Calculi which form but remain confined to the gallbladder are often
symptomless. However, if the stone migrates and in doing so becomes lodged in either the cystic or common bile ducts, a severe cramp like pain called biliary colic may be felt. Total blockage of the common bile duct usually results in obstructive jaundice, in which the skin acquires a characteristic yellow or olive green appearance, due to the retention of the bile pigment bilirubin in the blood. Faeces also lose their characteristic colour and appear either grey or white. If the condition persists chronic cholecystitis, infection of the bile ducts and cirrhosis of the liver may result.

The increased incidence of biliary stone disease in the industrialised countries of the west is contrasted only by its low incidence in less advanced countries such as Africa. Possibly the best indication of why this is so, may be seen in countries like Japan. Here the populous is increasingly adopting a western like, life style and more importantly a western style diet. In Japan\textsuperscript{42,43} not only has the incidence of gallstones increased from 1.7\% to 6.7\%, but at the same time a change in gallstone composition has occurred, cholesterol stones are now far more common than pigment stones. Many studies on the incidence of gallstones in the world have indicated that women\textsuperscript{44,45} are far more likely to develop gallstones than men. An autopsy study\textsuperscript{46} in nine British towns indicated that between 9.2\% and 20.6\% of those examined had gallstones. The highest figure of 20\% being associated with their occurrence in women and the lowest, for men in the same age group. Denmark and Italy also has a similar incidence to that in the UK. Other studies have shown that between 60\% and 80\% of cases are asymptomatic, i.e. only approximately 20\% of those with gallstone disease ever develop significant symptoms\textsuperscript{47-49}. Patients in the highest risk groups include those in older age groups, obese individuals undergoing weight reduction\textsuperscript{50,51}, those with hepatic cirrhosis\textsuperscript{52,53} and those with American Indian genes\textsuperscript{54}. Dietary factors can also have an important effect, Vegetarianism\textsuperscript{55} appears to reduce the risk of gallstones and animal studies suggests that alcohol may be protective\textsuperscript{56}.

More than 80\% of all gallstones reported in industrialised countries contain cholesterol\textsuperscript{57}. The amount of cholesterol present in the blood is dependent to a large extent on diet and the action of the liver. Under normal circumstances cholesterol is excreted by the liver as a component of hepatic bile. In cases where there is a considerable excess of cholesterol the bile can become supersaturated. It has been suggested that the first critical step in the formation of a cholesterol gallstone, is the nucleation and formation of micro precipitates of cholesterol from the supersaturated bile\textsuperscript{58}. These microprecipitates then form crystals of cholesterol monohydrate, which in turn aggregates to form macroscopic gallstones\textsuperscript{59}. However it has been shown that supersaturation of bile does not necessary result in cholesterol precipitation\textsuperscript{58} and
therefore other factors must play a roll in gallstone formation. One such factor is the presence of a gallbladder derived nucleating agent, the presence of the agent may also explain why there is a higher incidence of gallstones in the gallbladder than in the hepatic ducts.

Studies show that the agent responsible for the nucleation of cholesterol crystals is gallbladder mucus. It has been proposed that the hypersecretion of mucus glycoprotein occurs as a direct result of changes in gallbladder bile composition, specifically, the fatty acids of the biliary phospholipids, which accompanies the supersaturation of bile with cholesterol. This has been backed up by the results of studies on cholesterol fed animals, which indicate a link between hypersecretion of mucus glycoprotein and the nucleation of cholesterol crystals. Human studies have also indicated that the nucleation agent is present in some cases of supersaturated human bile (lithogenic bile) and that it accelerates cholesterol crystal formation. Mucin glycoprotein polymer may also assist nucleation by acting as a matrix or cement substance, such as an organic binder. Here the cholesterol-lecithin hydrophobic precipitates, may preferentially bind with region of the glycoprotein polymer and therefore initiate crystal growth. This has been seen in one study where cholesterol crystals were seen to form predominantly in the mucin packed niches of a human gallbladder. Another factor suspected of contributing to the formation of gallstones, is gallbladder motility, since it is known that motility is impaired in human patients with gallstones. In studies on guinea pigs on high cholesterol diets, gallbladder dysfunction has been shown to occur prior to the development of gallstones. It is likely that the mechanism which induces hypersecretion also effects gallbladder motility. Therefore in addition to the presence of high levels of cholesterol and nucleation defects, such as seen in patients with lithogenic bile, cholesterol crystal growth may also require a certain degree of gallbladder motor dysfunction. The so called triple defect. Cholesterol may also be precipitated if insufficient bile salts or lecithin are present in the bile.

2.42 Diagnosis

Gallstones are usually diagnosed on the basis of the patients reported symptoms. Asymptomatic stones are often only found during examinations for other complaints. Once diagnosed, more detail is often required about the stone, since the size, location, number and composition of the stone often determines the most suitable method of treatment. Some idea about the composition of a stone is often determined by the
characteristics it exhibits when imaged using the various visualisation techniques, available to the radiographer. One technique is radiography, most large, smooth, radiolucent stones are rich in cholesterol, however between 14% and 20% of those stones will also contain additional elements. Radiopaque stones are therefore excluded from peroral chemolysis, in which a number of bile acids are used to dissolve the cholesterol stones. This is despite the fact that approximately 33% of opaque stones are predominantly cholesterol. Another visualisation technique is oral cholecystography, here a substance given orally and opaque to X-rays, becomes concentrated in the gallbladder. Stones that exhibit buoyancy, are generally composed of cholesterol, however only 33% of all patients with cholesterol stones have buoyancy. Endoscopic retrograde cholangiopancreatography (ERCP), in which an endoscope is passed in to the duodenum and a contrast medium injected into the biliary duct and ultrasound techniques also yield important information about a stone. After the final diagnosis one of the following treatments will often be prescribed.

2.4.3 Treatment alternatives

As was the case with suprapubic lithotomy, the surgical treatment of gallstone disease prior to the introduction anaesthetic and antiseptics was invariably fatal. Soon after the introduction of antiseptic John Stough Bobbs, a self trained civil war surgeon from Indianapolis, Indiana, performed the first cholecystolithotomy. The operation was performed on the 15th of June 1867 on a 30 year old female and involved opening the abdomen and the removal of the stones through an incision in the gallbladder. For this landmark operation Dr Bobbs was later presented the American medical association award for distinguished service. The first elective cholecystolithotomy was performed by J. Marian Sims, a physician from Alabama in 1878. Four years later in 1882, a German physician called Karl Langenbuch described an alternative procedure in which the gallbladder was physically removed (cholecystectomy). He believed that this procedure was the least invasive of the two and that the gallbladder was physiologically irrelevant, since horses, deer and rats did not possess them. For the next 40 years, the two procedures were used side by side for the treatment of gallstone disease. Gradually however, cholecystectomy was adopted as the gold standard and is now the definitive treatment for symptomatic stones in the gallbladder. Cholecystectomy is currently the third most frequent operation in Germany and one of the commonest in Britain and the United States. Despite its widespread usage the approach does have its disadvantages, surgical mortality is approximately 0.5% overall and is higher among males and increases with age. Recuperation can also be
prolonged, the average hospital stay is generally around 8 days and it is normally another 20 days before full recovery is achieved.

One of the latest methods of removing the gallbladder which only requires minimal surgery is laparoscopic cholecystectomy. This effectively reduces the hospital stay to between 1 and 3 days, with full recovery being achieved in 1 to 2 weeks. On the downside however there is evidence that cholecystectomy may be a predisposing factor in the development of cancer of the ascending colon, particularly in women. The research suggests that an average 30% more cases of cancer of the colon and rectum occur after cholecystectomy than were expected in the general population. When the data was assessed by sex, the results indicated that astonishingly, women were almost 70% more likely to develop cancer of the colon after cholecystectomy than the general female population. Possible explanations for the increased rate of cancer among these patients are indicated in several studies. These suggest that after cholecystectomy, bile acid metabolism is altered and that its composition also changes. The metabolism changes occur due to the nearly continuous passage of bile through the liver and intestine, instead of only during digestion. This continuous recycling of bile results in increased exposure and degradation of the primary bile acids by intestinal bacteria, resulting in an increased proportion of secondary bile acids, such as lithocholic and deoxycholic acid. Further reaction may produce potent carcinogens. The increased incidence of right sided large bowel or ascending colon cancer, may therefore be associated with the higher levels of absorption of secondary bile acids in this first part of the large intestine.

Because of the complication associated with open surgery and the fear about cancer, alternative methods have been investigated. The last 10 years has seen several new minimally invasive procedures introduced, these are, percutaneous (endoscopic) cholecystolithotomy, methl tert-butyl ether lavage, peroral drug chemolysis and shockwave lithotripsy in association with peroral drug chemolysis. The details of these procedures will be discussed next.

2.431 Percutaneous (endoscopic) cholecystolithotomy (PCCL)

PCCL is one of a growing number of procedures in which the operation is conducted through the skin. In this particular procedure a fibre optic instrument called an endoscope is firstly introduced into the abdominal cavity and then into the gallbladder via several small incisions. The stone is then either physically removed or shattered
first using an ultrasound probe sent down the endoscope. PCCL\textsuperscript{91} is an endoscopic version of J Bobb's original cholecystolithotomy adapted from percutaneous nephrolithotomy\textsuperscript{92} a procedure successfully used in the treatment of kidney stones. Its major advantage over most of the other current procedures is its suitability for almost all types, sizes and numbers of gallstones. The ability to carry out bile composition tests is also particularly important, since this allows an evaluation of the possible cause and risk of recurrence of stones. This enables the necessary preventative measures, such as an adjustment to diet to be made. It also has the advantage of reducing the patients recovery time, to something similar to that experienced with laparoscopic cholecystectomy. Although, as with any surgical procedure there are risks from infection, haemorrhage and damage to the internal organs. However, follow-up studies of 39 of the first 60 patients\textsuperscript{93} has shown promising results.

2.432 Methl tert-butyl ether lavage

Methl tert-butyl ether\textsuperscript{94} (MTBE) is one of the latest cholesterol solvents to be used for the dissolution of cholesterol stones in the common bile duct and gallbladder. This agent is particularly useful because it has a high solvent capacity\textsuperscript{95} (14g/dl) and a boiling point of 55 °C. The solvent is administered percutaneously with a small calibre cannula by direct puncture of the gallbladder (transhepatic route). Gradual dissolution of the cholesterol gallstone is achieved by repeated instillation\textsuperscript{96} of MTBE into the gallbladder for several hours per day over a 3 to 4 day period. Occasionally during treatment the patient may experience biliary pain, nausea and sedation. Although more severe complications may arise if the solvent enters the liver or the blood stream via contact with a vascular structure. In such cases haemolysis, in which haemoglobin is lost from the red blood cells, renal failure and coma\textsuperscript{97} have been reported. Additionally in up to 5% of patients cholecystectomy will be required, due to persistent leakage from the puncture wound in the gallbladder. Patients with pigment stones, calcified stones or stones larger than 15mm are not suitable for MTBE treatment, although a combination of this and other treatments is currently under investigation. In the case of large stones, pretreatment with lithotripsy\textsuperscript{98} may be the answer. Dissolution of pigment stones with MTBE is not possible, although a new cocktail of dimethylsulphoxide 60%, MTBE 20% and sodium bicarbonate 20% has shown some promise\textsuperscript{99}. Generally patients treated with MTBE will also receive oral doses of bile acids (peroral chemolysis) for several months after treatment. In 34% to 70% of patients debris remains in the gallbladder, but eventually between 50% to 90% of patients\textsuperscript{96,100} will become stone free.
2.433 Drug (peroral) chemolysis

Dissolution of cholesterol gallstones by drug chemolysis is one of the main therapeutic alternatives to cholecystectomy. The administration of bile acids to patients was first described by Rewbridge in 1937. Further advances were made in the 70's when one particular bile acid component, chenodeoxycholic acid (CDCA), was found to reduce cholesterol saturation of bile and therefore cause the gradual dissolution of cholesterol gallstones. Early patient studies using CDCA over periods of 6 to 42 months, suggested stone free rates of 13% and 40% respectively for the two studies. However CDCA is known to have dose related side effects and cause perturbation of the serum levels of the liver enzymes and diarrhoea. Monotherapy with CDCA has now been largely superseded by another bile acid, ursodeoxycholic acid (UDCA) because of its lower toxicity and improved efficacy. Monotherapy with UDCA is unfortunately relatively expensive, costing in excess of $100 per patient per month. Therefore combinations of UDCA and CDCA are now commonly used. In patients with radiolucent stones (indicative of cholesterol) complete dissolution has been achieved in 60% of patients within 2 years. Those with buoyant, radiolucent gallstones ≤5mm in diameter have an even better chance, since 80% to 90% of patients are stone free within 1 year. The addition of terpenes (found in the essential oils of plants) to UDCA has also been found to improve results. There is also evidence to suggest that UDCA could be useful in the treatment of liver disease.

2.434 Shockwave lithotripsy plus drug chemolysis

The combination of extracorporeal shockwave biliary lithotripsy (ESBL) and orally administered bile acids is evolving rapidly as an effective new method for the treatment for gallstones. The method has improved to such an extent that patients are now treated on an outpatients basis i.e. general anaesthesia and an overnight hospital stay are no longer required. The shockwave lithotripsy component of the treatment has two major objectives. The first is to break the large or multiple stones into small fragments and thus improve the efficacy of bile acid dissolution therapy. Secondly to allow the spontaneous passage of fragments into the intestine. The reliance on the spontaneous passage of fragments may not be as an efficient mechanism for the elimination of the stone fragments as dissolution with UDCA and CDCA. This is due to the patients level of gallbladder dysfunction and the tortuous nature of the cystic duct with its valve of Heister. In general stone fragmentation is relatively straightforward and is approached in a similar way to that for renal stones,
although buoyant stones often make localisation difficult due to movement during treatment. The side effects of the treatment include soft tissue damage, hematuria, nausea, vomiting and biliary pain due to the passage of fragments. Animal studies also indicate that shockwave treatment causes moderate tissue damage to the liver and gallbladder, but that the damage is undetectable after 30 days. Severe side effects such as pancreatitis and cholecystitis occur only in 3% to 5% of patients and are mainly associated with the passage of stone fragments after lithotripsy. Those patients with single radiolucent stones with a diameter of ≤25mm have been found to be the most suited to this combined treatment. One of the latest studies has indicated that up to 70% of patients with single stones ≤20mm are stone free within 5 to 8 months of the start of the combined treatment. This is compared to 50% of patients with stones between 20mm to 30mm over the same period. However nearly 90% of all suitable patients with stones ≤30mm are totally stone free after 13 to 18 months. It can be seen that the disintegration of large stones is a crucial factor in the success of the treatment. Since it greatly shortens the time required with bile acid treatment to achieve a stone free state. Table 3 summarises the main features of the treatment alternatives.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Cholecystectomy</th>
<th>Drug chemotherapy</th>
<th>ESWL+Drug</th>
<th>MLBE</th>
<th>PCCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stone type</td>
<td>any</td>
<td>cholesterol</td>
<td>cholesterol</td>
<td>cholesterol</td>
<td>any</td>
</tr>
<tr>
<td>No. of stones</td>
<td>any</td>
<td>any</td>
<td>≤3</td>
<td>any</td>
<td>any</td>
</tr>
<tr>
<td>Stone size</td>
<td>any</td>
<td>≤10 mm</td>
<td>≤25 mm</td>
<td>≤20 mm</td>
<td>any</td>
</tr>
<tr>
<td>Ca in stone</td>
<td>ok</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>ok</td>
</tr>
<tr>
<td>Pigment stone</td>
<td>ok</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>ok</td>
</tr>
<tr>
<td>Treatment duration</td>
<td>1-2 hours</td>
<td>months-years</td>
<td>months</td>
<td>2-5 days</td>
<td>1 hour</td>
</tr>
<tr>
<td>Hospitalisation</td>
<td>8-10 days</td>
<td>out patient</td>
<td>out patient</td>
<td>2-5 days</td>
<td>1-3 days</td>
</tr>
<tr>
<td>Convalescence</td>
<td>2-3 weeks</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>1-2 weeks</td>
</tr>
<tr>
<td>Painful</td>
<td>yes</td>
<td>sometimes</td>
<td>sometimes</td>
<td>moderate</td>
<td>slight</td>
</tr>
<tr>
<td>Complications</td>
<td>many</td>
<td>diarrhea</td>
<td>obstruction</td>
<td>nausea</td>
<td>sepsicaemia</td>
</tr>
<tr>
<td>Prob stone free</td>
<td>~100%</td>
<td>~20%</td>
<td>~85%</td>
<td>~90%</td>
<td>~100%</td>
</tr>
</tbody>
</table>

Table 3 Biliary stone treatment alternatives.

### 2.44 Stone recurrence

A cholecystectomy can effectively remove the risk of stone recurrence in the gallbladder. But as with all the methods described, the possibility of recurrence in the bile ducts remains. All the forms of treatment in which the gallbladder remains in situ, generally run a greater risk of recurrence. In patients successfully treated with the bile acids UDCA and CDCA, it is estimated that up to 50% will have recurrent stones within a 3-5 year period. The stone recurrence rate for lithotripsy plus oral bile acids after 1 year is 9%, but may have stone recurrence rates similar to peroral chemolysis, because of the reliance on bile acids to dissolve the fragmented stone.
This is compared to 29% for MTBE\textsuperscript{119} over the same period, this increased incidence is thought due to the presence of flecks of material remaining in situ\textsuperscript{120} after treatment. A report on 667 patients undergoing surgical cholecystolithotomy\textsuperscript{121} between 1915 and 1937 indicates a recurrent incidence of between 14% - 30% over the average follow up period of approximately 20 years. The difference in recurrence, being associated with gender, type and number of stones. There was also uncertainty over whether the stones were actually recurrent or due to minute stones that had remained in situ after treatment. Patients that are probably least likely to suffer from a recurrence are those with single stones\textsuperscript{122} and adequate gallbladder motility.

Present methods for the prevention of recurrent stones in patients are based on either measures to lower cholesterol saturation of bile, or on attempts to influence cholesterol nucleation by controlling mucin formation\textsuperscript{46} and gallbladder motility. Manipulation of cholesterol saturation is presently achieved using intermediate doses of UDCA\textsuperscript{122}, intermittent therapy with high doses of UDCA and CDCA\textsuperscript{123}, or high fibre\textsuperscript{124}, low refined carbohydrate\textsuperscript{125} diets. There are however side effects associated with the use of bile acids, these include diarrhoea and liver damage. The findings of at least one study\textsuperscript{62} indicates that the control of nucleation by inhibiting gallbladder mucin glycoprotein secretion, may be a far more important way of preventing primary and recurrent gallstones. Gallbladder mucin glycoprotein secretion in cholesterol fed prairie dogs, has been shown to be inhibited using aspirin\textsuperscript{126} and other non steroidal anti inflammatory drugs. The reduction in both the quantity and concentration of the mucin in the aspirin treated animals, was shown to prevent the nucleation of cholesterol microprecipitates and thus gallstones, from the supersaturated bile. Long term trials with aspirin\textsuperscript{127} for the prevention of infarction, stroke or vascular death among patients with heart disease, has shown a 25% reduction in serious vascular events. Aspirin doses of 40 mg/day and larger single doses of 160 mg, appear to produce no significant adverse effects, apart from a few cases of gastrointestinal bleeding. Oral doses of aspirin used in conjunction with one of the treatment alternatives, described earlier, may therefore, prove a safe long term alternative for the prevention of gallstones in high risk patients.

2.5 Lithotripsy of renal and biliary calculi

Extracorporeal shock wave lithotripsy has now been routinely used for the treatment of renal calculi since 1982. The first investigation into the use of sound, specifically ultrasound, to disintegrate urinary and biliary calculi was conducted by Coats\textsuperscript{128} in
1948. Prior to this, work with high frequency sound waves, generated by quartz crystals had been undertaken in areas such as the underwater detection of submarines (Langevin 1917) and the detection of cracks in steel. Coats suspected that by subjecting a calculus to a beam of ultrasound the structure of the stone might become sufficiently weakened that it would disintegrate. In his experiments the stones were exposed to power densities of $5 \times 10^4$ W/m$^2$, for time intervals of 1 to 5 minutes at a range of frequencies between 100 and 1550 kHz. He showed that the best results were obtained with a frequency of 350 kHz at a power density of $5 \times 10^4$ W/m$^2$ for 5 minutes. But noted that the stones resisted disintegration and required an additional physical force (finger pressure) to crumble them. Stones subjected to higher frequencies required longer periods of treatment before the stone could be crumbled between the fingers. To determine the effect of the ultrasound on tissue Coats exposed a human kidney containing a stone to the above treatment. Examination of the kidney and the stone after treatment showed a similar softening of the stone and that no physical damage to the kidney had occurred. Although the temperature of the water in which the kidney was immersed was found to have risen to 65°C and blanched its outer surface. Exposure to energy densities of $18 \times 10^4$ W/m$^2$ for between 1 and 2 minutes was found to cause tissue damage and increase the water temperature to 78°C. Similar studies on tissue absorption indicated that the power density required to penetrate the intervening fat and muscle and still have an effect on a stone would result in severe tissue damage. Two German researchers did report the fracture of several types of biliary calculi with ultrasound, but it was necessary to expose the stones for more than 1 hour. A more direct method of delivering acoustic energy was suggested by Mulvaney. His applicators or buttons were made of glass, aluminium and steel and effectively collected the acoustic energy from the crystal and delivered via a tapered tool to the stone.

Because of the serious tissue damage which occurred when using continuous ultrasound at the intensity necessary to fracture calculi, the method was very soon rejected. The next and most significant advance came with an approach described by Häusler. The idea used acoustic shockwaves generated by an under water arc discharge. A patent application from Donier systems of Germany then followed and a year later Forbmann et al published a clinically applicable procedure. Following animal studies the first clinical application of shockwave lithotripsy to renal stones in 200 selected patients was undertaken between 1980-1982 using the Dornier lithotripter "the worlds most expensive bathtub". The use of shockwaves to fragment stones avoids the potentially serious and damaging effects of localised tissue heating experienced when using continuous ultrasound. This is in part due to the shockwaves
lower effective frequency which enables it to pass through fluid and body tissue with
less energy loss. The ultrasound fields used in lithotripsy are characterised by high
amplitudes, a long pulse duration and repetition rates of around 1 Hz. The low
repetition rate means that patients undergoing lithotripsy experience lower time
averaged ultrasound power density exposures, than even those receiving diagnostic
ultrasound. The widely cited threshold value is \(10^3\) W/m\(^2\), which is considerably
lower than the levels used by Coats. Below this value small amplitude continuous and
pulsed ultrasound has been found to produce negligible tissue heating\(^{137}\). Calculations
made by Filipcznski and Piechocki\(^{138}\), show that even with 5000 lithotripter pressure
pulses of 80 MPa, at repetition frequencies 100 times that used clinically, no more
than a 2°C rise in tissue temperature occurs.

2.51 Shockwave lithotripsy

A conventional lithotripter consists of two main components, a focused shockwave
source and an Xray or ultrasound localisation system for targeting the stones. The
shockwaves are generated by an emitter outside the body and transmitted as pulsed
longitudinal waves through a coupling medium and tissue to the stone. Shockwave
sources can be grouped into one of two categories depending on the characteristics of
the shockwaves produced. With supersonic emitters, shockwaves are produced by a
sudden release of energy, such as a plasma explosion in a small volume. The
electrohydraulic or spark gap source, laser and explosive pellet sources are examples
of this kind of emitter. Although presently only the spark gap source is commercially
used. Finite amplitude emitters rely on the electrical displacement or distortion of a
surface to produce an acoustic wavefront. Examples of this kind of device are the
electromagnetic source and the piezoelectric source, both of which have received
much interest in the last 10 years and which will be discussed next.

2.511 The electrohydraulic source

The electrohydraulic source utilises an underwater electrical discharge between a pair
of electrodes positioned at the geometric focus of an ellipsoid reflector\(^{139}\) figure 9a.
The discharge of a high voltage across the electrodes cause an explosive evaporation
of the water (plasma explosion), which initially expands with supersonic velocity\(^{140}\).
Because the initial expansion is greater than the speed of sound in the medium, its
motion is not immediatly detected by the surrounding fluid. Therefore, a spherical
wave of compressed fluid at a higher density, pressure and temperature than the surrounding fluid forms. This pressure wave or shockwave is characterised by a very steep change in pressure amplitude. The expansion of the plasma envelope finally stops when the pressure difference across the boundary of the cavity is equalised. As the cavity begins to collapse the negative phase of the wave is produced\textsuperscript{141}. The pressure wave is then reflected from the inner wall of the elliptic reflector and converges at the second focal point, significantly increasing the pressure. Peak pressures of around 100 MPa are typical. However, part of the wavefront propagates towards the second focal point without being focused. This direct component of the wave reaches the focal region in advance of the main component. The time difference between the two being determined by the focal length and dimensions of the reflector. Most modern sources remove the direct component of the wave by blocking it with part of the apparatus used to hold one of the electrodes.

The principle disadvantages of using this type of source is the high shock to shock variance in energy. This can be as much as 45\textsuperscript{\%}\textsuperscript{142} and is attributed to the inherent variability of the electrical discharge. There is also a minimum voltage necessary to generate a spark discharge at the electrodes. For clinical applications this is usually 14 kV, which corresponds to a pressure well above 50 MPa. This means that this type of source is not able to generate low pressures. With typical electrode lifetimes of approximately 4000 shockwaves, electrodes must be replaced after each treatment. This inconveniences operators and increases maintenance costs.

![Wave propagation](image)

Figure 9 (a) the electrohydraulic source (EH), (b) the piezoelectric source (PE), (c) the electromagnetic source (EM).

2.5.12 The laser source

Another recent addition to the growing armoury of treatment alternatives available for treating stone disease is the laser. Trials into the destruction efficiency of pulsed laser radiation from dye\textsuperscript{143}, excimer, Nd.YAG\textsuperscript{144} and several other sources on urinary and
biliary stones have been undertaken. Earlier studies used laser radiation to generate an explosive plasma at the focus of an ellipsoid reflecter\textsuperscript{145}. In one of the latest applications\textsuperscript{146}, excimer laser radiation is delivered to the stone surface through an endoscope, using a flexible silica glass optical fibre. On the surface power densities up to 1 GW/cm\textsuperscript{2} induce dielectric breakdown. This generates temperatures in excess of 10,000 K in the rapidly expanding plasma\textsuperscript{147}, resulting in the emission of a shockwave and the production of a cavitation bubble and liquid jets near the surface of the stone. Laser induced shockwave lithotripsy has recently undergone its first clinical trials and is now being used in patient tests in Germany. Using similar technology, laser sources are also being investigated for cleaning stonework and ceramics\textsuperscript{148} affected by air pollution.

2.513 The piezoelectric source

The use of piezoelectric shockwave sources in lithotripsy is a relatively recent application of piezo materials\textsuperscript{149}. Presently several manufacturers offer lithotripters using either a planar or spherical shockwave source. In both, large arrays of up to 3000 piezoceramic elements\textsuperscript{150} are electrically connected and arranged on a specially designed planar or spherical backing. The spherical arrangement of elements, similar to that shown in figure 9b, provides for the self focusing of shockwaves, whilst planar devices require an acoustic lens. Despite this additional requirement, planar devices can be built more compactly than spherical devices, which often have diameters of up to 50 cm. However, both types of source have the same principle of operation. When an electric field is applied across the array of piezoceramic elements, a displacement occurs due to the change in length of each of the elements, the piezoelectric effect. This collective displacement produces a pressure wave\textsuperscript{151}, immediately followed by a negative pressure wave as the array returns to its original shape.

In contrast to electrohydraulic sources, this type of device produces acoustic waves which, due to the non linear effects of water, distort during propagation to form a shockwave. The production of a single high intensity pressure wave is highly dependent on the impedence of the driving electronics and the form of the applied voltage pulse. This has to be matched to the electro-acoustical characteristics of the piezoceramic. Piezoceramic materials for use in lithotripters, such as PZT, have to be designed for high power and therefore require a high dielectric breakdown strength. Because of this, operating voltages are usually limited to between 2 kV and 6 kV, a factor of 10 below that routinely used with electrohydraulic and electromagnetic
sources. Therefore, despite high shockwave pressures (>100 MPa) due to the superior focusing ability of the large array, the amount of shockwave energy per pulse is small\textsuperscript{152}. This as we shall see later, appears to effect stone disintegration efficiency, leading to an increase in retreatment rate. The lifespan of the piezoelectric source is typically around 1,000,000 shockwaves. But is ultimately limited by mechanical damage and the loss of electrical insulation of the individual piezoceramic elements.

2.514 The electromagnetic source

The electromagnetic acoustic transducer (EMAT) is based on the principle of the electromagnetic shock tube\textsuperscript{153}. Its early uses included sonar sources for oceanographic research\textsuperscript{154} and the investigation of chemical equilibria. Development of the EMAT by Reichenberger and Naser\textsuperscript{155} of Siemens, has resulted in its widespread use as a shockwave source in lithotripsy. A schematic of the source is shown in figure 9c. The EMAT consists of a slab coil and a metallic membrane, separated by a thin insulating sheet. When a large discharge current is applied to the coil the membrane is displaced, producing a planar pressure wave in the adjacent medium. The lens then focuses the wavefront, which during propagation forms a shock wave, producing peak focal pressures of up to 60 MPa. The typical electro-acoustic efficiency of this type of device is around 0.03\% and therefore has the largest energy requirement of the sources so far discussed. The normal lifetime of the source is usually limited by metal fatigue in the metallic membrane. Although with an average maintenance period of between 200,000 - 400,000 shockwaves, up to 200 patients can normally be treated before the membrane needs be replaced. The major advantage of this type of device is its low variance in shockwave energy, typically less than 3\% over a large range of settings, allowing accurate control of treatment. Examples of typical operating parameters and the electro-acoustic efficiency for three of the main shockwave sources used in lithotripsy are shown in Table 4.

<table>
<thead>
<tr>
<th>Source</th>
<th>Operating voltage (kV)</th>
<th>Discharge current (kA)</th>
<th>Electrical energy (J)</th>
<th>Pressure P\textsubscript{R0} (MPa)</th>
<th>Pressure P\textsubscript{Rf} (MPa)</th>
<th>Efficiency (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electrohydraulic</td>
<td>10-30</td>
<td>10-30</td>
<td>15</td>
<td>21-78</td>
<td>3.6-9.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Piezoelectric</td>
<td>2-6</td>
<td>1-3</td>
<td>5</td>
<td>9-114</td>
<td>6.2-9.9</td>
<td>0.05</td>
</tr>
<tr>
<td>Electromagnetic</td>
<td>5-20</td>
<td>4-10</td>
<td>50</td>
<td>8-60</td>
<td>2-8</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 4 Summary of the typical operating parameters for the electrohydraulic, piezoelectric and electromagnetic sources.
2.52 Localisation systems

Stone localisation on many lithotripters is achieved using either Xray/fluoroscopy or ultrasound B-scan, although some of the latest generation of multifunction lithotripters combine both\(^{156}\). One of the many methods uses a bi-plane Xray system with image intensifiers. The two planes define a point in space or a virtual focus, which is also the focus of the shockwave source. For treatment the stone is simply identified on a screen, its coordinates calculated and the patient moved into position. Other methods employ inline Xray or ultrasound i.e. the system is integrated into the shockwave source in a similar way to that shown in figure 10. Those using inline ultrasound probes which only display 1 image plane have to be rotated to scan the whole volume. Once the stone has been located, the entire shockwave head is adjusted so that the stone lies within the target point indicated by the cross hairs on the monitor.

![Figure 10 Piezoelectric source with in line ultrasound imaging.](image)

One of the major advantages of using ultrasound, is that it avoids Xray exposure and permits online monitoring of the stone during treatment. This is particularly useful with gallstones\(^{157}\) since they often move in the course of treatment, making it necessary to continually correct the targeting. This type of system also allows the patient to control focusing with their breathing. This is far more effective than either computerised respiratory gating\(^{158}\), where the shockwaves are triggered in synchronism with the patients respiratory cycle, normally on expiration, or 'hit control\(^{159}\). It is also less expensive\(^{160}\) and suitable for the treatment of infants. All gallbladder stones can be visualised, although certain stones in the bile duct, urinary stones in the mid ureter, 20%-30% of other urinary stones and multiple stones generally require Xray visualisation. The inferior spatial resolution of ultrasound compared to Xrays also means that interpretation can be difficult, requiring longer user training. With Xray visualisation almost 85% of all kidney stones can be visualised, the remaining 15% and most gallstones require the use of a contrast medium (Xray
flouroscopy. X-ray visualisation techniques provide high contrast with good spatial resolution and can provide hard copy documentation (X-ray film) of the treatment. This means that the chances of missing stones is minimal. However worries about excessive radiation exposure prevents the on line monitoring of treatment. Figure 11 shows some typical X-ray images of a kidney stone.

![X-ray images of a kidney stone](image)

Figure 11 A sequence of X-ray photographs taken during ESWL treatment showing the gradual disintegration of a staghorn calculus in the right kidney.

### 2.53 Stone fragility

Urinary and biliary stones often respond very differently to fragmentation with shockwaves. In a recent study Dretler\textsuperscript{161} suggested that the capability of ESWL to fragment particular stones should be considered when selecting treatment, since its success appeared to depend on the stones fragility. Stone fragility has been shown to be largely dependent upon chemical composition,\textsuperscript{161} crystalline structure\textsuperscript{162} and flaw size. Microhardness,\textsuperscript{163} elasticity,\textsuperscript{164} stone volume and shape,\textsuperscript{165} gas content,\textsuperscript{166} as well as the properties of the fluid\textsuperscript{167} surrounding the stone also affect fragility. In the study calcium oxalate monohydrate (COM), calcium oxalate dihydrate (COD), sturrite, brushite, uric acid and cystine stones were subjected to 200 shockwaves generated by a Donier HM3 lithotripter. After treatment the percentage dry weight of each part of the calculi able to pass through a 2 mm sieve was calculated. From this a quantitative assessment of the fragility of each of the stones was made. The results showed that 100% of the uric acid and COD fragments were able to pass through a 2 mm grid and
therefore had the highest fragility. This is compared to only 64% of COM, 57% of sturvtie, 47% of brushite and 16% of cystine. In an attempt to make some sort of therapeutic distinction, the assessment for fragility was compared with measurements of linear density (table 1) and radiographic image density for each of the stones. The linear density was measured with a bone densitometer, using a technique called single photon absorbtionometry. Comparison of the results for linear density and radiographic image density indicated a reasonable level of correlation, but only a partial correlation with stone fragility. A comparison between microhardness, measured using a device known as an indenter and fragility, in a review by Coleman also showed little correlation. This is not altogether so surprising since the microhardness of kidney stones has been shown to be critically dependent upon urine pH. COM stones are significantly softer in alkaline urine than in normal urine. In addition it noted that the microhardness of most gallstones, typically 11-44 MPa, is an order of magnitude lower than that of kidney stones (190-650 MPa). Despite being softer they are typically more difficult to fragment. The presence of cholesterol in various quantities in the majority of gallstones appears to make no difference to fragility, although pigment stones with calcium carbonate shells resist fragmentation.

Studies of stone fragments resulting from ESWL have noted that fragmentation often occurs at the lines of circumferential or radial laminations and at the interfaces of crystals of differing composition. With sturvtie stones dislocation occurs between the sturvtie and calcium phosphate crystalline interfaces. So although sturvtie is soft and can be crumbled between the fingers, fragmentation results in a wide range of fragment sizes and thus, a low fragility. Cystine stones are known to have a uniform crystal structure, without laminations. Uric acid stones also have a uniform crystal structure but unlike cystine stone have very distinct concentric and radial laminations. Therefore, despite similar radiographic and linear densities, uric acid stones tend to be fragile whilst cystine are not.

2.54 Mechanisms of stone fragmentation

The exact mechanisms by which urinary and biliary stones are fragmented are still not completely understood. However, recent research with substitute stone materials such as chalk cubes, plaster of paris and synthetic stones have provided a better understanding of a few of the mechanisms. In general the formation of fracture planes and spalling are accepted as providing a significant contribution to the fragmentation process. Generally the larger and harder the stone is, the more
susceptible it is to spalling damage. Cavitation and liquid jets are also known to play a major role in fragmentation. In order to understand how a shockwave induces fracture and spalling damage, it might be helpful to look at the basic physics of a shockwave interaction with a stone.

When a wavefront is incident upon an interface, part of the wave is transmitted and part of it will be reflected. The fraction of the pressure amplitude transmitted or reflected is related to the acoustic impedance of the two materials, by the following standard equations. For simplicity it is assumed that the shockwave strikes the interface at normal incidence and that no attenuation occurs. In this instance then, the transmission coefficient \( P_t/P_i \) and the reflection coefficient \( P_r/P_i \) are given by

\[
P_t/P_i = 2Z_2/(Z_2 + Z_i) \quad P_r/P_i = (Z_2 - Z_i)/(Z_2 + Z_i)
\]

In which \( Z \) (the acoustic impedance) = \( \rho c \), and the subscript 1 refers to the first material, and 2 to the second, \( c \) is the velocity of sound in the medium and \( \rho \) its density. \( P_i \) is the incident pressure amplitude.

<table>
<thead>
<tr>
<th>Materials</th>
<th>Sound velocity (ms(^{-1}))</th>
<th>Density (kgm(^{-3}))</th>
<th>Acoustic impedance (10(^{-6}) kgm(^{-2})s(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air</td>
<td>344</td>
<td>1.293</td>
<td>0.0004</td>
</tr>
<tr>
<td>Water</td>
<td>1482</td>
<td>998</td>
<td>1.5</td>
</tr>
<tr>
<td>Muscle</td>
<td>1545-1630</td>
<td>1070</td>
<td>1.7</td>
</tr>
<tr>
<td>Fat</td>
<td>1460-1470</td>
<td>920</td>
<td>1.35</td>
</tr>
<tr>
<td>Kidney</td>
<td>1560</td>
<td>1040</td>
<td>1.62</td>
</tr>
<tr>
<td>Urinary stone</td>
<td>6260</td>
<td>1870</td>
<td>11.7</td>
</tr>
<tr>
<td>Gallstone</td>
<td>1400-2300</td>
<td>820-1100</td>
<td>1.15-2.42(^{174})</td>
</tr>
</tbody>
</table>

Table 5 Sound velocities, densities, and acoustic impedance's of various materials.

In this idealised example we shall consider a 60 MPa shockwave incident upon a gallstone with an impedance of 2 kgm\(^{-2}\)s\(^{-1}\), which is surrounded by tissue having an impedance of 1.5 kgm\(^{-2}\)s\(^{-1}\). This assumption for tissue impedance is sufficient, since the acoustic impedance of most body tissue is close to that of water. The relative acoustic impedances and properties of tissue and other materials can be found in table 5 and in the following reference\(^{175}\). A shockwave incident upon the first interface (water \( Z_1/\text{stone} \ Z_2 \) will be partially reflected, however the major proportion of the wave or 69 MPa will be transmitted \( (Z_1<Z_2 \text{ equation 2.01)\). The transmitted part of the wave propagates through the stone, where at the second interface, (\text{stone} \ Z_1/\text{water} \ Z_2 \) the stones rear surface, it will be partially reflected. Because the wave is moving from a high to a low impedance medium, the reflected part undergoes a phase change \( (Z_1>Z_2 \text{ equation 2.01)\) and becomes a tensile or rarefractional wave of -9.8 MPa, figure
12. As this wave propagates back into the stone it interacts with the remaining part of the wavefront, resulting in the tension AB. When the tensile strength of the stone is exceeded i.e. the stone is no longer be able to support the tension the material fractures. This usually results in the ejection of material from the rear surface, i.e. a spall.

![Diagram of tension and compression](image)

Figure 12 The production of fracture planes by tensile stresses.

The point at which the fracture plane develops in an homogenous material is defined by the duration and velocity of the shockwave. The thickness \( \delta \) of the spall or chip of stone is given by,

\[
\delta = \nu \Delta t
\]  

(2.02)

where \( \nu \) is the shockwave velocity and \( \Delta t \) the shockwave duration. Studies have indicated\(^{176}\) that typical kidney and gallstones are 7 times more sensitive to tensile stresses than they are to compressive stresses. The repeated exposure to these tensile stresses during shockwave treatment, results in the production of further fracture planes and the eventual fragmentation of the stone into progressively smaller parts. Fracture planes may also develop within the stone at the interface between laminations and liquid filled fissures. These abrupt changes in density can lead to the generation of tensile waves and therefore further aid the fracture process.

The effectiveness of a shockwave to fragment a stone can be severely limited by the presence of stone debris and sludge\(^{142}\). As debris builds up around the stone during treatment, urine or bile will permeate between the fragments creating additional interfaces. This effectively screens the stone by scattering a proportion of the incident shockwave. The reduction in energy reaching the stone can eventually halt the process of fragmentation. This may explain the difficulties incurred in treating impacted
ureteral stones. In these cases fragments are often prevented from moving away from the stone by a layer of tissue that has grown over the impacted stone.

2.541 Cavitation

The other mechanism known to contribute to the fragmentation of calculi is cavitation. Acoustic cavitation is brought about in this case by the rupture of the fluid by the rarefaction phase of the shockwave. If the negative pressure is low enough to reduce the local pressure to below the vapour pressure of the fluid, microscopic gas pockets called cavitation nuclei expand to form vapour filled cavities. All living systems are known to possess cavitation nuclei\textsuperscript{177} which are distributed in considerable numbers around the organism.

The threshold for acoustic cavitation in water is around -0.5 MPa, but depends on the surface tension, temperature, impurity and gas content of the water\textsuperscript{178,179}. As the shockwave passes by and the localised conditions change back to positive pressure, the cavity collapses. During this rapid collapse the potential energy gained in expansion is converted to kinetic energy\textsuperscript{180}. The concentration of this energy into a volume smaller than the original nuclei, creates temperatures and pressures as high as 10,000 K and 1000 MPa\textsuperscript{147}. In some cases these high temperatures and pressures result in the production of free radicals\textsuperscript{181} and sometimes in light, in a process called sonoluminescence\textsuperscript{182}. Normally most of the energy is radiated in the form of an intense shockwave, with peak pressures of around 500 GPa. However, if the bubble is near a boundary, the forced asymmetry in the fluid motion near the bubble, will cause it to collapse asymmetrically. In this case the kinetic energy from the collapse is converted into a liquid jet, which penetrates the interior of the cavity and impacts on the surface of the boundary\textsuperscript{183,184}. Lord Rayleigh\textsuperscript{185} was one of the first to explain that the erosion seen on ship propellers was due to cavitation. Liquid jet impact depressions and other damage associated with cavitation phenomena on thin sheets of aluminium and copper foil have been examined by Coleman et al\textsuperscript{186}. Their experiments have indicated that a single shockwave from a Dornier lithotripter (electrohydraulic type) can produce of the order of 25 jet impacts per cm\textsuperscript{2}. In some cases the jet impact was violent enough to puncture the thin foils. They suggest that it would be reasonable to assume that jet impacts also contribute to stone disintegration. Although the lower number of cavitation nuclei present in vivo, within the kidney for example, may reduce the number of impacts.
Evidence that cavitation plays an important role in the fragmentation of stones is provided by Sass\textsuperscript{187}. Analysis of gallstones after undergoing shockwave treatment showed the characteristic hallmarks of cavitation and of high speed liquid jet damage. Micro jet damage was also observed within small liquid filled fissures in the centre of the stone. High speed films of the shock interaction and electron micrographs of the resulting damage, demonstrate the importance of cavitation in the stone fragmentation process. It is suggested that cracks in the stone are initially produced by the tear and shear forces described earlier. Liquid be it bile or urine then penetrates these cracks resulting in cavitation activity and the eventual destruction of the calculus. Because bile is more viscous than urine\textsuperscript{188} the effectiveness of this mechanism to aid in the fragmentation of gallstones may be reduced. This is not so with urine which is able to penetrate cracks and fissures more easily. This may be one of the reason for the higher fragmentation rates seen with urinary stones.

2.542 Biological effects

The short term side effects of exposure to shockwaves from conventional lithotripters include haematuria, in almost all cases, hypertension, haematomas and pancreatitis. The passage of fragments cause renal and biliary colic, obstruction and urosepsis. Morphological changes in the kidney result in reduced output and fluid retention in the tissue and cavities of the body. The parenchymal trauma causes changes to the concentration of blood enzymes and proteins in the urine. In almost all cases these side effects are temporary, long term side effects such as diminished renal function have not been found. Slight tissue damage or bruising in the form of red and purple spots due to bleeding under the skin usually occur on the body at the shockwave entrance and exit sites. The reflection of shockwaves at other tissue-gas interfaces, such as the alveoli of the lung and at the intestine must be prevented. Studies of shockwave exposure on the thoracic region of rats i.e. the region of the trunk between the neck and the abdomen, resulted in extensive lung damage and death. Tissue has also been found to be damaged by pressures of around 40-50 MPa, when the shockwaves are focused on to the tissue itself\textsuperscript{189}.

A comparative study of tissue damage\textsuperscript{190} was conducted on renal tissue cells using the three main types of shockwave source. The results (table 6) for cell viability after exposure, indicated that the electromagnetic source, fitted to the Siemens lithostar caused the lowest amount of cell damage. Analysis of table 4 shows that this type of
device operates at the lowest pressure. With the particular model in the study, the maximum pressure available is 38 MPa.

<table>
<thead>
<tr>
<th>No. shockwaves</th>
<th>Dornier HM3 (Electrohydraulic)</th>
<th>Siemens Lithostar (Electromagnetic)</th>
<th>Wolf/Rezlith (Piezoelectric)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kV</td>
<td>2400</td>
<td>3000</td>
<td>4000</td>
</tr>
<tr>
<td>Viability %</td>
<td>66</td>
<td>19</td>
<td>40</td>
</tr>
</tbody>
</table>

Table 6 Cell viability after shock wave exposure.

The possibility of damage to cell DNA from shockwave exposure has also been investigated. In-vitro studies have shown that DNA damage does occur in regions of intense cavitation, due to localised heating or free radical generation. Clonogenic assays of these cells have indicated that it has no effect on the cell reproductive capacity. Although the possibility of more subtle non lethal changes has yet to be investigated. Presently no carcinogenic effects of shockwave exposure have been reported. Recent work in the area of shockwave exposure on bone has indicated that cavitation activity improves fracture healing, by stimulating new bone growth. There is now considerable interest in the use of ESWL for the treatment of conditions such as pseudo-arthrosis. Cell and vascular damage are also being investigated as a means of increasing the effectiveness of cancer therapy. The exposure of tumours to shockwaves has been shown to have little effect, apart from temporally slowing the growth of some tumours. There also seems to be some evidence that shockwave induced vascular damage in tissue around the periphery of a tumour can significantly slow the tumour growth rate. Vascular damage, which can result in cell death due to the reduction in blood supply, may also enhance hypothermia treatment of malignant tissue by reducing heat loss to the blood. There is also evidence of enhanced cytotoxic drug uptake resulting from changes in cell membrane permeability due to cavitation activity from shockwave treatment. Further details on the biological of shockwave exposure may be found in a review published by Brummer et al.

2.55 What makes a shockwave efficient in lithotripsy

An earlier discussion on tissue damage indicated that shockwave pressures above 50 MPa might be damaging to tissue. Other studies have shown that a minimum pressure of at least 20 MPa is necessary to fragment urinary stones. Taking account of shockwave attenuation due to tissue, approximately 15%, suggests a minimum requirement of 30 MPa. Ideally then, a lithotripter should produce pressures high enough to effectively disintegrate stones, but as low as possible to avoid severe tissue
damage. Clinical trials have shown that patient retreatment rates differ depending on the type of lithotripter used. For urinary stones, 20% of patients will require a second treatment session when electromagnetic or electrohydraulic lithotripters are used. When treated with a piezoelectric lithotripter, almost all require 2 or more sessions for stones larger than 15 mm. This suggests that there is another factor that affects the disintegration efficacy of lithotripters. In an attempt to resolve this, we should ask 'what makes a shockwave efficient in lithotripsy'. A recent publication by Granz and Köhler, with that very title, has attempted to address this question. In their experiments, craters were eroded in artificial kidney stones by shockwaves produced by an electromagnetic source. The amount of stone erosion was then correlated with a particular physical parameter of the shockwave. Their results indicated that there is a strong linear dependence ($r = 0.98$) between the stone erosion and the effective energy of the shockwave in the focus (figure 13a). Where the effective energy was defined as the energy density multiplied by the base area of the crater. Importantly, peak pressure ($r = 0.54$) figure 13b and rise time of the shockwave did not significantly correlate with eroded volume. This suggests that it is the total acoustic energy rather than peak pressure which is the dominant parameter and which affects the stone fragmentation efficiency of the lithotripters.

![Figure 13a](image)

**Figure 13a.** The dependence of the eroded volume on the effective energy and b, the dependence of the eroded volume on the peak pressure.

Knowing what makes a shockwave efficient enables us to analyse studies on the acoustic output of commercial lithotripters, such as the very detailed one by Coleman and Saunders or that by Folberth. The typical acoustic energy per pulse of an electrohydraulic source is the same order of magnitude as that from an electromagnetic source. However the limitations on minimum operating voltage, imposed by the discharge on the electrohydraulic source means that the electromagnetic source can operate effectively at much lower pressures, see table 4. This means that stone
fragmentation efficiency is high due to the high energy density and tissue damage is minimal. In contrast, the piezoelectric source requires high pressures and sharp focusing to achieve sufficient energy at the focus. Despite this, the energy per pulse is still low in comparison, resulting in lower fragmentation efficiency and higher retreatment rates. The very high pressures also indicate that tissue damage may be increased. Although the large aperture and high focusing gain means that peak pressures at the skin are low \( P < 0.6 \text{ MPa} \), \( (EH = 20\text{MPa}, EM = 6\text{MPa})^{201} \) enabling pain free treatment.

2.56 Lithotripter technology and future prospects

Since the introduction of the Dornier HM3, the first commercial lithotripter, ESWL has proven to be a world-wide success. The continual development of lithotripter technology has resulted in improved performance and the minimisation of side effects. The 'second generation' of lithotripters introduced in the mid 80's combined advanced technology with convenient shockwave generation and improved localisation systems. The overwhelming success of these systems has now led to the introduction of a 'third generation'. These interdisciplinary, multifunctional systems, can support both X-ray and ultrasound B-scan localisation and have a wide range of shockwave energies. There are currently upwards of 20 models available, table 7, offering various combinations of shockwave source, coupling and localisation systems. Some even offer painless, anaesthesia free treatment. With so many machines available, comparison is clearly very difficult. However, it is possible to identify some of the characteristics and features that would make the 'ideal' lithotripter. This machine would ideally combine ultrasound and X-ray localisation for urinary, biliary and salivary lithotripsy. It should also operate at pressures below which tissue damage occurs, especially for children and have a graduated shockwave energy range which is constant and reproducible. The peak shockwave energy at the focus should be as high as that obtained with electrohydraulic sources to provide the same degree of disintegration efficacy. Other essential requirements would be a large aperture and high focal gain for pain free treatment, simple operation and low maintenance costs.

Analysis of these requirements would suggest that it is the choice of the shockwave source that should be the main consideration. One source that appears to have many of these ideal characteristics and offers the required degree of flexibility is the electromagnetic source. Most of the experience with this type of source has been gained with the Siemens lithostar/lithostar plus\textsuperscript{203}. However, it is one of the new
generation and a recent variation of the electromagnetic source, the electromagnetic cylinder fitted to the Stortz modulith SL20\textsuperscript{20}, that appears to show the most promise. This design does away with the energy absorbing lens used with the conventional type of electromagnetic source, utilising instead a parabolic metal reflector. It is therefore capable of producing pressures and shockwave energy densities equivalent to the electrohydraulic sources fitted to the latest generation of Dornier machines. The design also allows the coaxial integration of an ultrasound probe enabling on line monitoring during treatment.

<table>
<thead>
<tr>
<th>Lithotripter</th>
<th>Source</th>
<th>Coupling</th>
<th>Localisation</th>
<th>Anaesthesia</th>
<th>Press/energy (M)</th>
<th>year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dornier HM3</td>
<td>EH</td>
<td>Water bath</td>
<td>2x Xray tubes</td>
<td>General</td>
<td>33-50 (0.09)</td>
<td>1980</td>
</tr>
<tr>
<td>low cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Direx X1</td>
<td>EH</td>
<td>Water cushion</td>
<td>Ultrasound / Xray*</td>
<td>Analgesia</td>
<td>n/a</td>
<td>1987</td>
</tr>
<tr>
<td>Northgate S03</td>
<td>EH</td>
<td>Water cushion</td>
<td>Lateral ultrasound</td>
<td>Analg / None</td>
<td>n/a</td>
<td>1987</td>
</tr>
<tr>
<td>Siemens lithostar</td>
<td>EM</td>
<td>Water cushion</td>
<td>Coaxial ultrasound</td>
<td>Analg / None</td>
<td>26-44 (0.017)</td>
<td>1989</td>
</tr>
<tr>
<td>ultras</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dornier comp</td>
<td>EM</td>
<td>Water cushion</td>
<td>Lateral ultrasound</td>
<td>n/a</td>
<td>n/a</td>
<td>1989</td>
</tr>
<tr>
<td>Stortz modulith</td>
<td>EMC</td>
<td>Water cushion</td>
<td>Ultrasound / Xray*</td>
<td>Analgesia</td>
<td>15-100</td>
<td>1991</td>
</tr>
<tr>
<td>SL5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Generation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dornier MPL 9000</td>
<td>EH</td>
<td>Water cushion</td>
<td>Coaxial &amp; lateral ultrasound</td>
<td>Analg / None</td>
<td>130</td>
<td>1987</td>
</tr>
<tr>
<td>Dornier MPL 5000</td>
<td>EH</td>
<td>Water cushion</td>
<td>Xray</td>
<td>strong</td>
<td>50</td>
<td>1988</td>
</tr>
<tr>
<td>Technomed sonolith 3000</td>
<td>EH</td>
<td>Partial water bath</td>
<td>Lateral ultrasound</td>
<td>Analg / None</td>
<td>52-78 (0.002)</td>
<td>1985</td>
</tr>
<tr>
<td>Siemens lithostar</td>
<td>EM</td>
<td>Water cushion</td>
<td>2x Xray tubes</td>
<td>Analg / None</td>
<td>26-44 (0.017)</td>
<td>1986</td>
</tr>
<tr>
<td>EDAP LT01</td>
<td>PE</td>
<td>Water cushion</td>
<td>Coaxial ultrasound</td>
<td>None</td>
<td>9-105 (0.003)</td>
<td>1986</td>
</tr>
<tr>
<td>Wolf piezolith 2300</td>
<td>PE</td>
<td>Partial water</td>
<td>2x in line ultrasound</td>
<td>None</td>
<td>56-114 (0.002)</td>
<td>1987</td>
</tr>
<tr>
<td></td>
<td>bath</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yachioda</td>
<td>LEAD</td>
<td>Water bath</td>
<td>Xray</td>
<td>Analgesia</td>
<td>n/a</td>
<td>1986</td>
</tr>
<tr>
<td>micropulsion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medstone 1050</td>
<td>EH</td>
<td>Water cushion</td>
<td>Xray</td>
<td>Analgesia</td>
<td>n/a</td>
<td>1987</td>
</tr>
<tr>
<td>3rd Generation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stortz modulith</td>
<td>EMC</td>
<td>Water cushion</td>
<td>Coaxial ultrasound</td>
<td>Analgesia</td>
<td>19-100</td>
<td>1989</td>
</tr>
<tr>
<td>SL20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siemens lithostar</td>
<td>EM</td>
<td>Water Cushion</td>
<td>Coaxial ultrasound</td>
<td>Analg / None</td>
<td>26-50 (0.053)</td>
<td>1989</td>
</tr>
<tr>
<td>plus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dornier MPL 9000-X</td>
<td>EH</td>
<td>Water cushion</td>
<td>Coaxial &amp; lateral ultrasound</td>
<td>Analgesia</td>
<td>130</td>
<td>1989</td>
</tr>
<tr>
<td>Dornier MPL 5000-U</td>
<td>EH</td>
<td>Water cushion</td>
<td>Lateral ultrasound + Xray</td>
<td>Analgesia</td>
<td>90</td>
<td>1990</td>
</tr>
<tr>
<td>Wolf piezolith 2500</td>
<td>PE</td>
<td>Water cushion</td>
<td>In line ultrasound + in line Xray</td>
<td>None</td>
<td>100</td>
<td>1989</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDAP LT02</td>
<td>PE</td>
<td>Water cushion</td>
<td>Coaxial ultra / Xray</td>
<td>None</td>
<td>100</td>
<td>1991</td>
</tr>
<tr>
<td>Dioracic</td>
<td>PE</td>
<td>Water cushion</td>
<td>Coaxial ultra + Xray</td>
<td>None</td>
<td>n/a</td>
<td>1988</td>
</tr>
</tbody>
</table>

Table 7 Current commercially available lithotripter systems.

* Requires time consuming interchange of ultrasound or Xray probes.

A comparison of the disintegrative efficacy of six machines has been undertaken by Rassweiler et al.\textsuperscript{204} In the study, the eroded volume of craters produced in plaster cubes by each of the machines, operating at the extremes of their range were compared after 100 shockwaves. The results, figure 14, clearly indicated the superior efficacy of the Storz modulith, which produced a crater with a volume of 450 mm\textsuperscript{3} after only 50 shockwaves. However, the Siemens lithostar and the Dornier machines are all very capable lithotripters and continue to treat hundreds of patients a year. As our knowledge of the process of fragmentation and of stone recognition improve, we might yet see a fourth generation of totally automated lithotripter.

![Figure 14 The disintegrative efficacy of six different lithotripters.](image)

Undoubtedly ESWL has a bright future, it has become a valuable asset to the urologist and greatly benefited those suffering from stone disease. During the short period since its introduction the management of stone disease has changed completely. To date world wide over 700 lithotripters have been installed, resulting in the treatment of over 1 million patients. Lithotripsy has now almost completely replaced open surgery in the treatment of urinary stones. Further research has now opened up new and exciting uses, such as the treatment of gallbladder and common bile duct stones. Clinically the most interesting is the treatment of pancreatic duct and salivary duct stones, where in the case of salivary stones, surgical complications such as partial paralysis of facial nerves can be avoided. Other possible applications actively being researched, are its use in the treatment of patients suffering from pseudoarthrosis of fractures. Research is also well advanced into the cytotoxic and sensitising effects of shockwaves on tumour tissues, implicating a possible use in the treatment of cancer. Today's treatment represents a far cry from the early days of the lithotomist. One wonders, what changes in history may have occurred if a few of the famous victims of the stone, had the benefit of lithotripsy.
Chapter 2: A history of the stone

References


1a Adams, F. The genuine works of Hippocrates translated from the Greek. Baltimore, Williams and Wilkins Co, 1938.


5 Cited by Hirsch Geographical and historical pathology

6 Bryant, A. Samuel Pepys, the man in the making. Cambridge University Press 1933.


37 Thompson, H. practical lithotomy and lithotrity; or an inquiry into the best modes of removing stone from the bladder. London, John Churchill and sons, 1863.


Chapter 2: A history of the stone

Chapter 2: A history of the stone


Chapter 2: A history of the stone


115 Tint, G.S. Salen, G. Chazen, D. Symptomatic gallstones are likely to reocur after dissolution with ursodeoxycholic acid (UCDA) but this may be prevented by low dose UCDA (Abstract). Gastroenterology. 92:1787, 1987.


145 DOS 25 38 960: Einrichtung zum berührungsfreien zertrümmern von im körper eines lebewesens befindlichen konkrementen.


149 Riedinger, R. 1986.


Chapter 2: A history of the stone


Chapter 2: A history of the stone


CHAPTER 3

THE GENERATION AND FOCUSING OF ACOUSTIC SHOCK WAVES

3.1 Introduction

Electromagnetic (EM) type devices, such as the electromagnetic acoustic transducer (EMAT) have already been identified as versatile, effective and reliable sources of shock waves. Factors found to be particularly important in lithotripsy for the treatment of stone disease. With lithotripsy currently the standard treatment for urinary stones and its rapid acceptance as a treatment for biliary stones, the use of the EMAT and its later generation cousin, the electromagnetic cylinder (EMC) is likely to become widespread. The EMAT belongs to a group of sources known as finite amplitude emitters. When activated they produce acoustic waves, which during propagation are distorted by the non-linear characteristic of the medium into the sawtooth form of a shock wave. The propagation of these finite amplitude waves and the formation of a shock front in this way can be described by a non-linear differential equation called the wave equation. A naturally occurring example of non-linear distortion can be seen in the propagation and breaking of waves on a beach. In its unfocused form, the EMAT produces waves of relatively low amplitudes, consequently a shock like structure only forms at considerable distances from the source. To obtain the operating pressures required for lithotripsy, focusing devices such as acoustic lenses are required. With focusing, a substantial increase in pressure and energy density and a reduction in the shock formation distance can be obtained. Typically, pressures of 50-100 MPa, energy densities of 0.5 mJ mm\(^2\), and rise times of tens of nanoseconds are produced in the focal regions of these devices. However, with acoustic lenses aberrations and diffraction effects can be considerable, so care must be taken when designing such lenses. Similar care has to be taken in the design of measurement equipment such as hydrophones. The rapidly changing pressures of an acoustic shock wave call for hydrophones with fast response times and good dynamic strength. The latter is particularly important, since the life time of a hydrophone in the focal region can be quite short. The importance of an accurate and reliable hydrophone is seen in the need to precisely determine the properties of the acoustic fields generated by these shock wave sources, for both theoretical and practical reasons.
3.2 Acoustic theory

3.21 Acoustic radiation

Acoustic radiation appears as a time varying change of pressure which travels progressively through an acoustic medium. This pressure change, which is equal to the difference between the ambient and the instantaneous values is known as the excess or acoustic pressure. Acoustic waves fall into two categories, homogenous and inhomogenous waves. Rayleigh and Love waves, both examples of inhomogenous waves propagate in only one plane and are only usually found on the surface layer of an acoustic medium. Their major usage is in seismological studies, where low acoustic absorption due to their low frequency allows them to propagate over considerable distances. The two types of homogenous wave are the transverse or shear wave and the longitudinal wave. Transverse waves set up large shear stresses within the medium, only solids and certain liquids with high viscosities can sustain these stresses without shearing. Therefore they cannot propagate through gaseous or the majority of fluid media. By far the most important wave in acoustics and the one that will be considered from now on is the longitudinal or compression wave, a diagrammatic model of which is shown in figure 1.

![Diagram of the longitudinal wave]

Figure 1 The longitudinal wave.

During propagation, particles in successive regions of the medium, along the path of the wave are displaced, accumulating at some points and becoming rarefied at others. As a particular region of the wave moves on, a restoring force returns the particles to their undisturbed positions. The accumulation or rarefaction of particles appears as a continuous change in density. The relation between density, indicated by the length and colour of the bars and the pressure is shown in figure 1. Regions of high pressure (compression) correspond to points of high population density. Similarly, low pressure
regions (rarefaction) occur at points with a low density of particles. Particle velocity, not to be confused with the propagation velocity, is greatest in regions where the extremes of pressure occur and where particle displacement is zero. This reduces to zero at points of maximum displacement or where the pressure is zero. Particles in the compressed regions move forward in the direction of propagation and those in the rarefied regions move backwards. Since successive particles within the medium along the path of the wave are displaced, the acoustic media must be compressible. This means that a wave cannot be transmitted instantaneously through the medium, but will propagate with a finite velocity, which is dependent upon the compressibility and density of the medium.

### 3.22 Acoustic velocity

The velocity $c$ with which a sound wave travels in a medium such as water is determined by the strength of the forces among the molecules. These forces are characterised by the bulk modulus $K$ and is a measure of how hard it is to compress the substance. When the pressure on an object is increased its volume decreases, resulting in an increase in density $\rho$. The bulk modulus relates the fractional density change $\Delta \rho/\rho$ to the pressure change $\Delta P$.

$$\Delta P = K \frac{\Delta \rho}{\rho} \quad (3.01)$$

Materials such as air are easy to compress, hence $K$ is small, stiff materials such as steel are very hard to compress and therefore have large bulk moduli. The velocity $c_o$ of a wave with an infinitesimally small pressure amplitude is found to depend only on the bulk modulus, which can be assumed to be constant and the density of the medium. The velocity $c_o$ is given by,

$$c_o = \sqrt{(K/\rho)} \quad (3.02)$$

where $K$ is the adiabatic bulk modulus. However, as the pressure amplitude is increased the volumetric strain fails to increase in proportion i.e. the medium becomes progressively more difficult to compress, resulting in a considerable increase in the value of the bulk modulus above its low amplitude value. A number of studies have examined the variation in compressibility$^1$ and the velocity$^{2,3}$ of sound in water as a function of temperature and pressure. The earlier velocity measurements made by Wilson$^2(1959)$ are in good agreement ($\pm 0.2$ ms$^{-1}$) with those of the later, more
Chapter 3: The generation and focusing of acoustic shock waves

accurate study by Barlow and Yazgan \(^3\) (1967). However, because Wilson’s results cover a wider range of temperatures and pressures, they will be cited in all further discussions. Wilson’s results were obtained by measuring the transit time of a short sound pulse at various temperatures and pressures, through a known liquid path length. These were then analysed using the method of least squares to obtain empirical equations for the speed of sound as a function of temperature and pressure.

\[
c = a_0 + a_1T + a_2T^2 + a_3T^3 + a_4T^4
\]  

(3.03)

where \(a_i = \sum_{j=0}^{3} (b_j)j \ P^j\)  

(3.04)

Here \(T\) is the temperature (°C) and \(P\), the absolute pressure (psia). The coefficients \((b_j)\) are given in the cited text. Using equations 3.03 & 3.04, the velocity of sound as a function of pressure (converted to Pascals) was calculated for a range of constant temperatures and then plotted in the graph of figure 2.

![Graph showing velocity of sound in distilled water as a function of pressure.](image)

Figure 2 The velocity of sound in distilled water as a function of pressure.
3.23 The shock wave

The wave motion of an acoustic wave as it propagates through water can be described by the wave equation. The wave equation relates the variations in pressure along various directions in space as seen at a particular instant of time, to the variation with time which occurs at any particular point in space. If the medium is continuous and homogenous and the alternations in pressure are so rapid that no appreciable heat energy can flow i.e. the process can be considered adiabatic and the variations in pressure are small compared to the average pressure, the three-dimensional acoustic wave equation can be written thus.

\[
\frac{\partial^2 P}{\partial t^2} = \frac{K}{\rho_0} \nabla^2 P \quad \text{where} \quad \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (3.05)
\]

If the acoustic pressure is independent of both the \( y \) & \( z \) axis i.e. it acts only in the \( x \)-direction equation 3.05 reduces to the simple one dimensional expression describing the propagation of plane waves.

\[
\frac{\partial^2 P}{\partial t^2} = c_0^2 \frac{\partial^2 P}{\partial x^2} \quad (3.06)
\]

The assumption that the wave amplitude was infinitesimally small means that the correspondence between the pressure and density of the medium could be considered linear i.e. the bulk modulus is independent of the acoustic pressure. For waves of finite amplitude the non linear terms of the wave equation are no longer negligible, due to the material non linearity and the effects of convection. The effect of material non linearity is seen in the relationship between pressure and density. As the pressure is increased, this relationship becomes non linear, as a result the bulk modulus increases causing the local speed of sound in the medium to increase, or decrease in the case of negative pressure. The convective non linearity of the medium is associated with the particle velocity \( u \), which varies throughout the wave as shown in figure 1. The greater the local acoustic pressure the larger the particle velocity in that part of the wave. Since the particle velocity in regions of compression are in the same direction as the phase velocity \( c \), these regions will tend to propagate at the velocity \( c + u \) and migrate faster than the regions of rarefaction, where \( u \) is in the opposite direction to \( c \). Therefore instead of the velocity of propagation being a constant, it actually varies and depends on the form and the amplitude of the acoustic wave. Hence for finite
amplitude waves, equation 3.06 must be modified to include the effects of the material and convective non linearity. The local speed of sound $c$ is now given by,

$$c = c_0 + \left(1 + \frac{B}{2A}\right)u$$

(3.07)

here the unity term corresponds to the contribution from the convective non linearity and the other term to the material non linearity. $B/A$ is the second order non linearity ratio of the medium, for water at 20 °C, $B/A = 5^5$. A finite amplitude wave propagating through a medium therefore becomes progressively more distorted, since the velocity in the higher pressure regions of the wave is greater than that in the rarefied region. Consequently the parts of the wave in the high pressure regions gain continuously on those in the rarefied regions, resulting in the wave becoming steeper in front and more gradual behind. These non linear effects can ultimately lead to the production of a highly unharmonic shock wave, characterised by an extremely steep change in pressure amplitude, i.e. a shock front. The transition of a finite amplitude acoustic wave into a shock wave is shown in figure 3. The individual waveforms show pressure as a function of distance.

![Figure 3 The nonlinear propagation of a finite amplitude wave.](image)

After propagating a certain distance (the discontinuity length) the initially sinusoidal waveform becomes saw toothed in appearance, taking the appearance of a shock wave. The rapid changes in particle density, velocity and temperature, produce the dissipating processes of diffusion, viscosity and thermal conduction. These compete against the effects of non linearity to produce a stable front, this situation persists until no more energy is delivered from the wave crest to maintain the slope. Severe attenuation then sets in reducing the amplitude of the shock wave until the ageing shock wave approaches a low amplitude sinusoidal form.

The characteristic parameters of a shock wave are shown in figure 3. $P_m$ and $P_r$ are the peak positive and peak negative pressure (MPa) respectively. The shock front rise time
t, is defined as \((t_{90\%} - t_{10\%})\), which in lithotripsy is typically between 10..600 ns. \(t_1-t_0\) and \(t_2-t_1\) define the duration of the compression and rarefaction phase of the shock wave respectively and \(t_w\) is the pulse half width time \((t_{1/2})\). The discontinuity length or distance \(S\) from the source at which the front becomes discontinuous and the waveform resembles a shock wave can be estimated for a plane wave by\(^4\).\(^5\)

\[
S = \frac{3 \rho_0 c_0 t_{1/2}^3}{\pi P_m} \left(2 + \frac{B}{A}\right)^{-1}
\]

(3.08)

where \(\rho_0\) and \(c_0\) are the density and speed of sound in the medium at normal pressures, \(P_m\) is the peak pressure and \(t_{1/2}\) is the half width of the acoustic pulse (1/3 of the period of the initial sine shaped pressure wave). Equation 3.08 shows that by increasing the pressure amplitude \(P_m\) and thus the speed differential between the peaks and troughs of the wave, the shock wave can be made to form closer to the source. However, this is subject to an upper limit due to the increased dissipation and eventual acoustic saturation\(^6\). In a focused system, the distance \(S_f\) from the lens at which a shock wave forms can be estimated by\(^5\),

\[
S_f = F(1 - \exp\left(-\frac{S}{F}\right))
\]

(3.09)

where \(F\) is the focal length of the lens.

### 3.24 Specific acoustic impedance

The impedance can be defined as the ratio of a general driving force to the velocity response. In acoustics, the driving force is the acoustic pressure amplitude \(P\) and the velocity response, the particle velocity \(u\). The acoustic impedance \(Z\) is therefore given by

\[
Z = \frac{P}{u}
\]

(3.10)

The impedance presented by a medium to an acoustic wave is not only dependent upon the properties of the medium but is also a function of the mode of propagation. Generally \(P\) and \(u\) are not in phase and thus the impedance \(Z\) is complex. However, for this analysis we will only consider plane waves in which \(P\) and \(u\) are in phase. From
the general solution for the simple harmonic form of the 1-D wave equation, the acoustic pressure and particle velocity are given by.

\[ P = P' \exp(i(\omega t - kx)) \quad P' = \rho_0 c u \]

and

\[ u = u' \exp(i(\omega t - kx)) \]

where \( P' \) is the peak acoustic pressure and \( u' \) is the peak particle velocity, \( \rho_0 \) is the density and \( c \) the speed of sound in the medium. The specific acoustic impedance \( Z \) from 3.10 & 3.11 for a plane wave propagating in an infinite medium is thus,

\[ Z = \rho_0 c \quad \text{kgm}^{-2}\text{s}^{-1} \]  

(3.12)

If the medium is not infinite, but is bounded by a plane surface normal to the direction of propagation, standing or stationary waves are produced due to interference between the direct and reflected waves. Under these conditions the acoustic impedance is no longer constant but alternates, having different values at different points along the wave. The velocity, density, acoustic impedance and attenuation of several materials commonly used in acoustics are shown in table 1.

<table>
<thead>
<tr>
<th>Material</th>
<th>Density kg m(^{-3})</th>
<th>Velocity m s(^{-1})</th>
<th>Acoustical Impedance (10^6 \text{kg m s}^{-2})</th>
<th>Attenuation N m(^2) (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air*</td>
<td>1.293</td>
<td>344</td>
<td>0.0004</td>
<td>1.8x10(^{-11}) (1x10(^3))</td>
</tr>
<tr>
<td>Water*</td>
<td>998</td>
<td>1482</td>
<td>1.479</td>
<td>25x10(^{-15}) (1)</td>
</tr>
<tr>
<td>Polystyrene(^{#})</td>
<td>1060</td>
<td>2350</td>
<td>2.5</td>
<td>23 (2.5x10(^{4}))</td>
</tr>
<tr>
<td>Perspex(^{#})</td>
<td>1185</td>
<td>2750</td>
<td>3.25</td>
<td>57 (2.5x10(^{4}))</td>
</tr>
<tr>
<td>Aluminium</td>
<td>2700</td>
<td>6320</td>
<td>17</td>
<td>0.4 (10x10(^{4}))</td>
</tr>
</tbody>
</table>

Table 1 The acoustic properties of some common materials. * @ 20°C.

A comparison between the attenuation coefficients for the different materials in table 1, shows that in the case of water this is extremely small, its effect will therefore be neglected in any later calculations.

### 3.25 Reflection and transmission of acoustic waves

Analysis of a wave in an infinite medium is only possible theoretically because in practice every medium terminates somewhere, i.e. it has a boundary. If the wavelength of the sound can be considered small compared to the dimensions of a boundary, between two materials on which it is incident then it will undergo reflection and
Chapter 3: The generation and focusing of acoustic shock waves

refraction in the same way as that described by geometrical law for light. It will also obey the same basic laws at the boundary, i.e.

a) The angle of incidence is equal to the angle of reflection, and

b) The incident ray, the reflected ray and the normal to the plane of incidence all lie in the same plane.

If the sound wavelength is comparable to the linear dimensions of the boundary the geometrical laws no longer apply and the phenomena is essentially due to diffraction. Figure 4 shows the incident, reflected and transmitted components of an acoustic wave at a plane boundary between two materials with acoustic impedances of \( Z_1 \) and \( Z_2 \).

![Figure 4 Plane waves incident upon a plane boundary.](image)

In considering the problem two boundary conditions must be met. Firstly, that the normal component of the particle velocity must be continuous across the boundary if the two materials are to remain in contact. Secondly that the acoustic pressure is also continuous across the boundary, otherwise the boundary would experience a net force. Expressing these conditions mathematically we have,

\[ u_i \cos \theta_1 + u_r \cos \theta_1 = u_c \cos \theta_2 \]  
(3.13)

and

\[ P_i + P_r = P_t \]  
(3.14)

since for plane waves \( P = \rho_0 cu \) (Equ. 3.11), 3.14 can be rewritten thus,

\[ c_1 \rho_1 u_i - c_1 \rho_1 u_r = c_2 \rho_2 u_t \]  
(3.15)

where the negative sign indicates the direction of the reflected wave. Eliminating either \( u_i \) or \( u_r \) from equations 3.13 & 3.15 we obtain the reflection coefficient \( R \) and the transmission coefficient \( T \).
\[ R = \frac{u_r}{u_i} = \frac{\rho_2 c_2 \cos \theta_1 - \rho_1 c_1 \cos \theta_2}{\rho_2 c_2 \cos \theta_1 + \rho_1 c_1 \cos \theta_2} \] \hspace{1cm} (3.16)

\[ T = \frac{u_t}{u_i} = \frac{2 \rho_2 c_2 \cos \theta_1}{\rho_2 c_2 \cos \theta_1 + \rho_1 c_1 \cos \theta_2} \] \hspace{1cm} (3.17)

The reflection and transmission coefficients apply to the particle amplitude, particle velocity, particle acceleration and acoustic pressure. Equations 3.16 & 3.17 can be replaced by their general form (Equ. 2.01) if the waves strike the surface at normal incidence i.e. when \( \theta_1 = \theta_2 = 0 \). When the incident and reflected intensities are to be compared the acoustic power reflection coefficient \( a_r \) is used,

\[ a_r = \frac{I_r}{I_i} = \left( \frac{\rho_2 c_2 \cos \theta_1 - \rho_1 c_1 \cos \theta_2}{\rho_2 c_2 \cos \theta_1 + \rho_1 c_1 \cos \theta_2} \right)^2 \] \hspace{1cm} (3.18)

similarly, the acoustic power transmission coefficient \( a_t \) is given by.

\[ a_t = 1 - a_r = \frac{4\rho_1 c_1 \rho_2 c_2 \cos \theta_1 \cos \theta_2}{(\rho_2 c_2 \cos \theta_1 + \rho_1 c_1 \cos \theta_2)^2} \] \hspace{1cm} (3.19)

In addition to reflection and refraction of the wave, mode conversion often occurs at the interface. As a result obliquely incident longitudinal or transverse waves may generate both longitudinal and transverse components in both the reflected and refracted waves.

### 3.3 The electromagnetic acoustic transducer (EMAT) operation

The electromagnetic acoustic transducer or EMAT shown in figure 5 is a large diameter, high energy acoustic emitter. The device produces acoustic waves in a similar manner to that of a piston acting on a liquid, except here the piston is replaced by a thin metallic membrane or diaphragm. Its construction is similar to a device described by Eisenmenger\(^9\), which was later adapted by Reichenberger and Naser\(^10\) for use in the Siemens Lithostar lithotripter. The device consists of a spiral slab coil, mounted on the face of an acoustically hard, electrically insulating, backing. Opposing this is a metallic membrane separated from the coil by an insulating sheet. The whole assembly is clamped into the mounting flange to form what is essentially an air gap transformer with a shorted secondary.
Chapter 3: The generation and focusing of acoustic shock waves

3.20 The Lorentz force $F$ is the force acting on a charge $q$ in an electric field $E$ or the force experienced by a charge moving with a velocity $v$ in a magnetic field of induction $B$. The Lorentz force is given by,

$$F = q(E + v \times B)$$

where the product $qE$ is the electric field force $F_E$ acting on a charge and $qv \times B$, the magnetic field force $F_B$. The absence of an electric field means that the contribution of the electric field force can be neglected. Therefore only the magnetic field force $F_B$ is associated with the displacement of the membrane. The total force is thus,

$$F = g \times B$$

where $g$, the induced current density, is the product of $qv$ and the resultant force $F$ is the cross product of $g$ and $B$ i.e. the force is mutually perpendicular to both the applied field $B$ and the induced current $g$.

The device is operated by discharging a large current from a capacitor, through the flat spiral coil. The highly damped current $i$, produces a very strong rapidly changing magnetic field $H$ around the conductor of the coil. The resulting localised magnetic flux $B$ induces an emf in the membrane, which act around closed paths causing very large induced currents, or eddy currents of density $g$ to flow. However, this is not brought about by the relative movement of the membrane, or the source of the field, as is normally the case in say, electric motors, but as a direct result of the strong time
varying magnetic field. From *Lenz's* law these mutually induced eddy currents, must flow in such a direction as to oppose the change in the magnetic flux and thus the coil current $i$ that caused them. A simplified model of the coil its magnetic field $H$, the induced current $g$ and the local magnetic flux $B$ acting on a unit volume $dV$ of the membrane is shown in figure 6.

![Figure 6](image)

Figure 6 The force experienced by a metallic membrane in a time varying magnetic field.

As a result of the current flowing through the EMAT coil, a force proportional to $i^2$ will be experienced by the membrane. This force, as indicated in equation 3.21, will be mutually perpendicular to both the induced current $g$ and local magnetic field $B$. Its direction can be found using Flemings left hand rule, where the first finger represents the direction of the field, the second the induced current and the thumb, the direction of the resultant force. Using this rule, it can be seen that the membrane undergoes a displacement away from the coil and that this is so, even when the driving current is reversed. This means that the membrane oscillates at twice the frequency of the driving current. The magnitude of the force on the membrane depends on the numerical value of the product $Bg$ and because $B$ and $g$ increase with $i$, the driving current should ideally be as large as possible. The distance between the coil and the membrane is also important, as is the resistivity of the membrane material. The separation should be minimised inorder to maximise the flux linkage and the membrane resistivity as low as possible to maximise the induced eddy currents.

### 3.3.1 The capacitor discharge system

The capacitor discharge system shown in figure 7 in the form of a simplified circuit diagram, provides the high transient current necessary to operate the EMAT. The circuit consists of a $0.544 \mu F$ high voltage capacitor, which can be charged and then discharged via a spark gap switch through the EMAT. Electrically the EMAT has an inductive impedance and is thus represented in the diagram by the inductance $L_e$ and its series resistance $r$. Connected in parallel with the EMAT is the protection resistance.
\( R_p \), which provides protection in the event of the EMAT coil going open circuit. The high resistance of \( R_p \), means that it has little or no effect on the circuit under normal operating conditions. Protection is also provided in the form of a separate earth for the high voltage side of the system. This is taken back to a substation earth grid, to prevent local electrical interference from voltage spikes in the mains. The spark gap, which consists of a trigger electrode and two hemispherical brass discharge electrodes separated by an air gap, is connected in series with the storage capacitor and the EMAT to form a series LCR circuit. Its trigger electrode which provides the means of initiating current flow in the circuit, is connected to the HV winding of an isolated step up transformer. Attached to the low voltage side of the transformer are the various timing and fire control circuits, which also control the laser schlieren system (chapter 4).

During the charging cycle, in which the capacitor is charged from the high voltage source through the resistance \( R \), the spark gap is non conducting. This results in a large potential difference, equal to the voltage across the capacitor and very close to the static breakdown voltage of the air gap, existing between the discharge electrodes. On completion of the charging cycle a trigger pulse is applied to the high voltage transformer. This initiates the discharge cycle in which a high voltage spark of approximately 20 kV pre ionises the gap between the trigger and the earthed discharge electrode. The main air gap then quickly breaks down due to the strong potential gradient that exists between the discharge electrodes to form an arc discharge. The near zero resistance characteristic of this arc, rapidly discharges the capacitor through the EMAT which emits a strong acoustic transient into the surrounding water. The strength of the transient is dependent upon the discharge energy. This can be changed by making the necessary adjustment to the distance between the spark gap electrodes prior to changing the voltage. The fixed operating parameters of some of the components in the circuit limit the operating range to between 15-20 kV.

![Figure 7 The EMAT driving circuit.](image-url)
3.311 The spark gap

Triggered spark gap switches are a very effective way of switching large currents. They are able to change extremely rapidly from a state of near perfect insulation, to that of a low impedance conductor through which large currents can flow and then recover their original insulating state. Typical operating parameters of a spark discharge include, currents of thousands of amperes, an arc resistance of milliohms and inductance's of between 5-30 nH. The spark gap switch used here is operated at atmospheric pressure at a voltage very close to its static breakdown voltage. This ensures that the shot to shot variation in delay time or jitter associated with the breakdown of the air gap is minimised.

The delay is attributed to the time lag immediately after the spark gap is triggered, before a suitably placed electron appears in the gap to initiate a breakdown in which an appreciable current flows. An electron or positive ion thus placed, will, under the action of the applied field be accelerated towards either the positive electrode (anode) in the case of the electron, or the cathode for the positive ion. In doing so they will inevitably suffer one of two types of collision in which energy will be lost to other electrons, ions and neutral gas atoms. In elastic collisions the transfer of energy is related to the mass ratios of the particles. This means that electrons suffer very little energy loss in this type of collision with ions and atoms. The electrons ability to maintain its velocity allows it to contribute to ionisation and excitation by inelastic collision. The one initial electron will therefore start an electron avalanche, resulting in a conducting channel being established across the gap. As the current builds up and the voltage across the gap collapses, energy supplied to the channel create gas temperatures of up to 50,000°K. The rate of increase of the current and the period of the discharge are ultimately controlled by the external circuit, after which the gap returns to its non conducting state.

3.312 Charging cycle

In the charging mode of operation the driving circuit of figure 7 can be reduced to the equivalent circuit of figure 8. This consists of the current limiting resister R the storage capacitor C and the high voltage source V.
Chapter 3: The generation and focusing of acoustic shock waves

Figure 8 The equivalent circuit in charging mode.

For convenience it will be assumed that the charging switch is closed at \( t = 0 \) and that the excitation voltage \( V_i \) is a step function defined by

\[
V_i(t) = \begin{cases} 
0 & \text{when } t < 0 \\ 
V & \text{when } t \geq 0 
\end{cases}
\]

Applying Kirchoff voltage law to the equivalent circuit of figure 8 for \( t > 0 \) gives.

\[
V - V_R - V_C = 0 \tag{3.22}
\]

Substituting the voltage/current relationships for a resistor and a capacitor gives us

\[
V = iR + \frac{1}{C} \int i \, dt \tag{3.23}
\]

The solution of this first order differential equation can be obtained by differentiating both sides, separating the variables and integrating with respect to time. The current \( i(t) \) at \( t = 0 \) is given by,

\[
i(t) = \frac{V}{R} \exp(-\frac{t}{RC}) \tag{3.24}
\]

where the current decays exponentially from its initial value at \( t = 0 \) to zero at \( t = \infty \). Similarly, the voltage \( V_c \) across the capacitor is given by,

\[
V_c(t) = V(1 - \exp(-\frac{t}{RC})) \tag{3.25}
\]

where the voltage increases exponentially at \( t = 0 \) to \( V \) at \( t = \infty \). The product \( RC \) is the time constant \( \tau \) of the circuit. The total energy \( E \) stored in the capacitor is,
\[ E = \frac{1}{2} CV^2 \]  

(3.26)

i.e. for a given capacitance the energy is proportional to the square of the voltage.

### 3.313 Discharge cycle

Once the capacitor has been charged to its working potential \( V_c \), the system is ready to be fired. The equivalent LCR series circuit in discharge mode is shown in figure 9.

The circuit consists of the storage capacitor \( C \) charged to a value \( q_0 \) (C.V) coulombs, the circuit resistance \( R_c \) which combines the spark gap switch resistance and the wire resistance and the EMAT coil represented by an inductance and its series resistance. Applying Kirchoff voltage law to the equivalent circuit of figure 9 and letting \( R = R_c + R_{\text{coil}} \) we have

\[ L \frac{di}{dt} + IR + \frac{1}{C} \int idt = 0 \]  

(3.27)

The differential equation for the circuit can be obtained by differentiating equation 3.27 with respect to time.

\[ L \frac{d^2i}{dt^2} + Ri + \frac{i}{C} = 0 \]  

(3.28)

The solution, obtained by solving the auxiliary equation, is of the form,

\[ i = A \exp(-\alpha + \beta)t + B \exp(-\alpha - \beta)t. \]  

(3.29)
where $A$ and $B$ are arbitrary constants, depending on the conditions of the circuit at the time of closing the switch. Assuming that at $t = 0$, $i = 0$ and that the capacitor voltage is $V$, then the current $i$ in the circuit on discharge is given by,

$$i = \frac{V}{\omega L} \exp\left(-\frac{Rt}{2L}\right) \sin \omega t$$

where the angular frequency $\omega$ is given by

$$\omega = \sqrt{\left(\frac{R}{2L}\right)^2 - \frac{1}{LC}}$$

If $\left(\frac{R}{2L}\right)^2 < \frac{1}{LC}$ then the system will act as a damped oscillator. The relationship between the coil current (Equ. 3.30) and the force on the diaphragm (chapter 3.3), for a coil resistance $R$ of 0.036$\Omega$ an inductance $L = 1.67$ $\mu$H and a discharge capacitance of 0.544 $\mu$F is shown in figure 10.

![Figure 10](image_url)

Figure 10 The relation between the coil voltage, current and the force on the diaphragm.

**3.32 The EMAT sound field**

Transducers of finite size, with the exception of perfectly spherical and point transducers, with an entirely radial surface motion, always have some degree of directionality. This effectively means that the sound waves will be confined to a beam
or solid angle over which they will be effective. The degree of directionality will be
effected by factors such as the transducers size and shape, the radiation impedance and
the mode of vibration of its surface. In an earlier discussion the operation of the
EMAT was compared to that of a circular piston acting on a liquid. Whilst this
comparison has its problems, especially where the resulting sound field is concerned,
sufficient similarity exists for us to use the analogy initially. The flat piston can be
thought of as being composed of a number point emitters arranged over its entire
surface area. If each of the emitters behave as simple harmonic oscillators and vibrate
in phase, with the same amplitude and frequency. Then, the radiated sound field is
similar to that observed for a plane wave incident upon an extended boundary in which
a circular aperture or diaphragm of the same diameter is present. The Huygen-Fresnel
principle\(^\text{13}\) shows us that in the central zone of the piston or diaphragm, the spherical
waves produced by the point emitters form plane wave surfaces. But owing to
diffraction at the edge (edge effects) annular waves are produced. Superposition of all
these wavelets (considering their amplitude and relative phases) produce in the near
field or Fresnel region, a sound field characterised by extreme interference in which
maxima and minima values of the acoustic pressure occur. The ratio of the radiator
diameter \(D\) to the sound wavelength \(\lambda\) determines the number of these maxima and
minima and also the spread of the interference field. This can be calculated from
diffraction theory, where the divergence angle \(\theta\) is given by,

\[
\sin \theta = 1.2 \frac{\lambda}{D} \quad (3.32)
\]

where \(\theta\) is the angle with respect to the normal of the radiator. The total beamwidth is
simply \(2\theta\). By inspection as the diameter to wavelength ratio increases, so the angular
divergence of the beam decreases. The complex structure of the sound field produced
by diffraction, can be made more uniform if the excitation of radiator is permitted to
decrease towards the edge or rim. In contrast to the piston source, the rim of the
EMAT is less strongly excited than the centre. Thus the axial pressure fluctuations that
characterise the near field of a piston source are flattened out with the EMAT.

Figure 11 The excitation and acoustic pressure along the axis of (a) a piston source
and (b) a radiator with a Gaussian excitation.
For a circular radiator\textsuperscript{14}, with a Gaussian like excitation, such as the EMAT, these fluctuations disappear almost completely as shown in figure 11. The extent of the near field $N$ for such a radiator is given by,

$$N = \frac{R_o^2}{\lambda} \quad \text{m} \quad (3.33)$$

where $R_o$ is the radius for which the excitation decreases to $1/e = 0.37$ of its maximum value. Excitation in the form of short wave trains or pulses can also greatly influence the sound field. Pulses coming from different points on the radiator may be unable to produce interference because they do not arrive simultaneously. Therefore, this also has the effect of making the sound field more uniform. In addition to the axial differences between the two sources, diffraction also results in the production of null points and side or minor lobes. These are often seen in the directional characteristics\textsuperscript{15}, which show the pressure distribution across the beam width in the plane ($z$ axis) normal to the transducer or $x$ axis. For the piston source with uniform excitation, the sound pressure profile in the far field is described by the directivity factor $R$,

$$R = \left( \frac{2J_1(\beta r \sin \theta)}{\beta r \sin \theta} \right) \quad (3.34)$$

where $J_1$ is the Bessel function\textsuperscript{16}, $\beta = 2\pi/\lambda$ and $r$ is the radius of the radiator. Equation 3.34 indicates that for certain angles from the axis the sound pressure is equal to zero, whilst at others it will reach a maximum. However, for transducers similar to the EMAT, with a Gaussian excitation, these additional maxima are not present and only a single beam with a gaussian like distribution across its width is produced\textsuperscript{17}. The far field acoustic pressure distribution across the beam width is shown for a piston source of radius $r$ and the EMAT ($R_o = r$) in figure 12.

![Figure 12 The directivity function for (a) the piston source and (b) the EMAT.](image-url)
Because of the absence of null points on the directional characteristics of sources such as the EMAT, the divergence angle can no longer be defined by equation 3.32. Instead it is defined by the angle made by the part of the characteristic with the axis at which the pressure drops to 70% (-3 dB) of its axial value.

3.33 Estimation of shock wave pressure and energy

When a large current flows through the coil of the EMAT a force is exerted on the diaphragm. This repulsive force displaces the diaphragm resulting in the emission of a planar acoustic pulse into the surrounding water. An estimate for the value of the pressure maximum, \( P_m \), of the pulse in the vicinity of the diaphragm has been made by Eisenmenger\(^9\) based on the peak discharge current \( I_m \).

\[
P_m = 2\mu_0 \frac{n^2 I_m^2}{D^2} \quad \text{Pa} \quad \text{where} \quad I_m = \frac{V}{\omega L} \quad \text{(Equ.3.30)} \quad (3.35)
\]

Also, \( D \) is the coil diameter, \( n \) the number of turns and \( \mu_0 \) the permeability of free space. Equation 3.35 shows that for a coil of \( n \) turns with a diameter \( D \), the peak pressure is proportional to \( I_m^2 \). Another parameter of the shock wave and the one identified in the study by Granz (chapter 2.55) as effecting the stone fragmentation efficiency of lithotripters, is the total acoustic energy or the energy per pulse. To calculate this it is first necessary to determine both the temporal and axial profile of the shock wave. Recent studies by Church\(^18\) and that of Philipp\(^19\) have shown that the variation of shock wave pressure \( P(t) \) with respect to time, can be modelled as the voltage response of a parallel RLC circuit to a current impulse. By defining the output in terms of the pressure, \( P(t) \) is given by,

\[
P(t) = 2P_m e^{-\alpha t} \cos(\omega t + \frac{\pi}{3})K(1-e^{-\beta t}) \quad \text{pascals} \quad (3.36)
\]

where \( \alpha \) is the decay constant and \( \omega = 2\pi f \) the radial frequency. The factor \( K(1-e^{-\beta t}) \) allows the selection of non zero rise times, e.g. a 100 ns rise time can be selected by choosing \( \beta = 4.6 \times 10^7 \text{ s}^{-1} \). The constant \( K = 1.03 \). Substituting these values and those for a shock wave with a peak pressure of 20 MPa, a period of 6 \( \mu \)s (\( \omega = 1.0467 \times 10^6 \text{ rads} \text{s}^{-1} \)) and a decay constant of (for a highly damped response) \( \alpha = 9.1 \times 10^5 \text{ s}^{-1} \), into equation 3.35 gives the pressure waveform shown in figure 13.
The intensity $I$ of an acoustic wave is defined as the rate at which energy is transferred across a unit area normal to the direction of propagation of the wave. From theory the instantaneous intensity $I$ at a point $r$ on the radius of a plane wave is given by,

$$I = \frac{1}{\rho c} \int_{t_1}^{t_2} P^2(r,t) dt$$

(3.38)

where $t_2 - t_1$ is the pulse length. Since energy = intensity x area, the total energy $E$ per pulse contained within a circular area defined by the radius $r_N$ is,

$$E = \frac{1}{\rho c} \sum_{n=1}^{N} \int_{t_1}^{t_2} P^2(r_n,t) dt \cdot A_n$$

(3.39)

where $\rho$, $c$, and $P$ have their usual meanings, $r$ is the radius and the area $A_n = (r_{n+1}^2 - r_n^2)\pi$. The total energy can therefore be calculated by measuring the wave pressure as a function of time at $N$ different radii from the $x$ axis. An alternative option, by which an approximate value for the shock wave intensity $I$ and total energy $E$ is obtained can be achieved by substituting the expression for $P(t)$ (Equ. 3.36) into equation 3.38 and then integrating over the period of the shock wave. This rather complicated integral (Appendix 1) was solved using Maple.
3.4 Acoustic focusing

Sound waves incident upon a boundary between two media undergo reflection, refraction and scattering in the same way as light. Because sound has these properties it can also be focused. Methods of focusing\textsuperscript{21} range from reflection systems, which employ spherical and parabolic reflectors, to refraction systems such as prisms and acoustic lenses. In addition, zone plates, similar to optical Fresnel - zone plates and waveguides make up a third group known as interference systems. However, it is the acoustic lens that finds the most application in ultrasonics. Generally the design of an acoustic lens is approached in a similar way to that of conventional optics. Provided several practical points are first taken into account, acoustic lenses can be made from both solids and liquids. In addition to those materials listed in table 1, lithium and carbon tetrachloride\textsuperscript{22} have also been used. Apart from its ability to maintain the complicated form of a lens, a suitable material should also (a) have a similar specific acoustic impedance as the surrounding medium in order to reduce, if not eliminate reflections at the interfaces; (b) a different acoustic velocity from the surrounding medium, to ensure good focusing power; (c) as low an absorption as possible and (d) a good degree of homogeneity to prevent scattering.

With liquid lenses reflection losses are negligible, typically < 1% of the incident energy is reflected, absorption is also low. However, the liquid must be contained in a thin walled enclosure that moulds the liquid into the shape of the lens. Enclosures of soft plastic or thin rubber help to reduce losses, but if a ridged lens is required thick walls are necessary, which lead to greater losses. Solid lenses made from plastics such as perspex and polystyrene (table 1) have a good impedance match with water, resulting in reflection losses of 12-16%, but an absorption that increases with frequency. Metals on the other hand have low absorption but a very poor impedance match, for aluminium, the reflection losses are around 70%. In addition to increased losses, solid lenses are also subject to mode conversion at the surface of the lens. This can lead to the propagation of both longitudinal and transverse waves, which due to their different propagation velocities are refracted at different angles, resulting in the production of a second focus\textsuperscript{23}.

Acoustic lenses also suffer from the effects of diffraction, which tends to be more evident than in conventional optics. This is a direct result of the wavelength to diameter ratio of the lens. Generally, because the lens is not infinite in extent only part of the incident wavefront will be accepted and some will be diffracted, thus the lens is said to be diffraction limited which will ultimately limit its imaging capability. Good
resolution depends loosely on making the ratio of the wavelength $\lambda$ of the radiation to the diameter $D$ of the lens as small as possible. By comparison, for light a system with a 136 mm diameter lens and $\lambda = 600$ nm, has a wavelength to diameter ratio of $4.4 \times 10^{-6}$. An acoustic system with the same lens, a frequency of 166 kHz and a sound velocity of 2350 ms$^{-1}$ has a $\lambda/D$ ratio of $1.4 \times 10^{-2}$. The value for the acoustic system is 4 orders of magnitude higher than that for light. Spherical aberration also becomes a problem in acoustic lenses but this can be eliminated by giving the lens surface an aspherical form, or reduced by the correct choice of surface radii, the best form approach.

3.41 Lens design theory

A lens is an optical system consisting of two or more refracting interfaces. From theory a ray incident upon an interface will be refracted according to Snell's law. If the rays considered are restricted to paraxial rays i.e. only those that make a small angle $\theta$ with the optical axis, such that $\sin \theta = \theta$ (first order approximation), then the focal length of a thin lens is given by the lensmaker's or thin lens equation. Recalling the lensmaker's or the thin lens equation from geometrical optics and adopting the standard sign conventions (appendix II) we have,

$$\frac{1}{f} = \frac{1}{s} + \frac{1}{s'} = (n_{lm} - 1)(\frac{1}{r_1} - \frac{1}{r_2})$$

where $f$ is the paraxial focal length, $s$ the object distance, $s'$ the image distance and $r_{1,2}$ the radius of curvature of the first and second lens surfaces respectively. The refractive index $n$, as in conventional optics, is the ratio of the speed of propagation in free space, in this case water, to that of the lens material. The acoustic velocities of several materials used in acoustics lenses are given in table I. From these the refractive index of water ($n_m$) and a polystyrene lens ($n_l$) are found to be 1 and 0.6306 respectively. The refractive index of the lens relative to the water $n_{lm}$ is thus 0.6306. Unlike conventional optics, where the value of $n$ is always greater than unity, in acoustics $n$ may be either greater or less than unity. A value less than unity results in the refractive index term of equation 3.40 becoming negative. Therefore, a sound wave incident upon a concave lens actually converges to a focus, the exact opposite to what would be expected for light incident upon the same lens in a conventional situation. The thin lens equation is only suitable for cases where the lens thickness is considered small compared to its other optical parameters, such as its focal length, object and image distances. When the lens thickness cannot be considered small, the lens must be
treated as a thick lens and a modified form of the lens makers equation used. The standard form has once again been modified to take account of the underwater environment and the effective focal length w.r.t. the principle planes is now given by,

\[
\frac{1}{f} = \left( \frac{n_l - n_m}{n_m} \right) \frac{1}{r_1} - \frac{1}{r_2} + \frac{(n_l - n_m)d}{nr_1r_2},
\]

(3.41)

where \(d\) is the thickness of the lens and all other parameters and sign conventions are as conventional optics.

### 3.4.11 Aberrations

The analysis in the previous section was restricted to rays that arrive at shallow angles in an extremely narrow region about the optical axis. These paraxial rays form the basis of first order theory or Gaussian optics. In practice these rays constitute only a small fraction of the total number of effective rays entering a system. Thus first order theory can be considered no more than a good approximation for obtaining the focal point of a beam of rays incident upon a lens. Departures from these idealised conditions of Gaussian optics are known as aberrations, of which there are two main types. Monochromatic aberrations, also known as Seidel aberrations consist of spherical aberration, coma, astigmatism, field of curvature and distortion, the other type is chromatic aberration. If a satisfactory image is to be obtained, it is necessary to reduce the effects of aberrations to a minimum. Acoustic systems employing large diameter lenses are particularly prone to spherical aberration. Here oblique rays parallel to the axis and incident upon a lens are not brought to a focus at a unique point. A reasonably accurate method of representing this deviation from first order theory is third order theory, in which the first two terms of the Maclaurin expansion of \(\sin \theta\) are used instead of just the first (first order theory). As a result an additional term which varies as \(h^2\) is introduced in to the expression for calculating the conjugate points of a single refracting interface. Rays striking the interface at greater distances \(h\) above the axis are focused nearer the vertex. The distance between the axial intersection of an oblique ray (a height \(h\) above the axis) and the paraxial focal point is termed the longitudinal spherical aberration (LSA).

The amount of spherical aberration for a fixed aperture and focal length will vary depending upon the object distance and the shape of the lens. It can be completely eliminated for a single lens by giving the surfaces an aspherical form. Unfortunately this is relatively expensive and therefore common practice is to adhere to simple
spherical surfaces and the reduction of spherical aberration by the correct choice of radii. This is achieved relatively easily by insuring that the incident ray makes more or less the same angle as does the emerging ray. Hence for a particular set of system parameters, f, s, s' etc there is a lens shape for which spherical aberration is a minimum. The lens shape can be expressed using the Coddington shape factor q.

\[ q = \frac{r_2 + r_1}{r_2 - r_1} \quad (3.42) \]

The system parameters can be expressed by the position factor \( P \),

\[ P = \frac{s' - s}{s' + s} = 2f \frac{s - 1}{s'} = 1 - \frac{2f}{s'} \quad (3.43) \]

where \( s \) and \( s' \) are the object and image distances respectively and \( f \) is the focal length. From theory the relation between the shape and position factor to produce minimum spherical aberration is given by,

\[ q = \frac{-2(n^2 - 1)P}{n + 2} \quad (3.44) \]

Substitution of \( s \), \( s' \) and \( r_1 \), \( r_2 \) from equation 3.42 & 3.43 into the lensmakers formula, gives the radii of the two surfaces in terms of \( q \) and \( f \).

\[ r_1 = \frac{2f(n - 1)}{q + 1} \quad (3.45) \]

and

\[ r_2 = \frac{2f(n - 1)}{q - 1} \quad (3.46) \]

Dividing 3.45 by 3.46 gives the ratio of the two radii.

\[ \frac{r_1}{r_2} = \frac{(q - 1)}{(q + 1)} \quad (3.47) \]

A negative ratio indicates that the surfaces curve in opposite directions.
3.412 Ray tracing

By applying Snell's law to the refraction at each interface of a thick lens, it is possible to trace the paths of several representative paraxial and oblique rays through a system. This method of determining the aberration is known as ray tracing. Initially each interface is taken in isolation and the angles of incidence and refraction calculated using Snell's law and basic trigonometry. At the first interface shown in figure 14 we have.

\[ \sin \phi_1 = \frac{h}{r} \quad (3.48) \]

where \( h \) is the height above the axis of the incident ray and \( r \) the radius of curvature of the first surface. From Snell's law the angle of refraction \( \phi_1' \) is related to the angle of incidence \( \phi_1 \) by,

\[ n \sin \phi_1 = n_1 \sin \phi_1' \quad (3.49) \]

where \( n \) and \( n_1 \) are the refractive indices of the surrounding water and the lens material respectively. Since the sum of all the angles in the triangle OBM are equal to 180°, \( \theta' \) is given by

\[ \theta' = \phi_1' - \phi_1 \quad (3.50) \]

By applying the sine rule to the triangle OBM the image distance \( s_1' \) can be found

\[ s_1' = r - \frac{r \sin \phi_1'}{\sin \theta'} \quad (3.51) \]
The image produced at $s_1'$ by the first interface now becomes the object or virtual object for the second. At the second interface figure 15 we have

![Diagram](image.png)

Figure 15 The second lens interface separated from the first by the distance $d$.

The object distance $s_2'$ for the second interface is given by,

$$s_2' = d - s_1'$$

(3.52)

where $d$ is the distance between the two interfaces i.e. the lens thickness. By applying the sine rule to triangle OLM, the angle of incidence $\phi_2$ made by the ray from the virtual object at M with the normal to the second interface can be obtained.

$$\phi_2 = \sin^{-1}\left(\frac{r_2 + s_2'}{r_2}\right)\sin\theta'$$

(3.53)

Once again from Snell's law the angle of refraction $\phi_2'$ is given by

$$\sin\phi_2' = \frac{n}{n_1}\sin\phi_2$$

(3.54)

and since all the interior angles of triangle MLM' must equal 180°

$$\theta'' = \phi_2' + \theta' - \phi_2$$

(3.55)

The final image distance $s_2''$ for the thick lens can therefore be obtained by considering triangle OLM'. From the sine rule $s_2''$ is given by

$$s_2'' = r_2 - \frac{r_2 \sin\phi_2'}{\sin\theta''}$$

(3.56)
3.42 The focal region

An earlier discussion indicated that the focal region of the EMAT-lens combination would be greatly affected by both spherical aberration and diffraction. In the case of spherical aberration, this leads to the real focus being extended along the transducer axis (x axis), whilst diffraction will result in the spreading of the field normal to the axis (z axis). The Gaussian like excitation of the EMAT diaphragm also means that interference and points of zero intensity, common in the field of a uniform emitter (figure 12), will be absent. As a result the focal region can no longer be defined by the point at which the pressure distribution across the field drops to zero. Instead it must be defined by the position at which the pressure drops to a specific value, this is commonly 50% of the maximum focal pressure (-6 dB zone). An estimate of the focal geometry and pressure gain for a radiator with harmonic excitation in the linear limiting case is provided by Rosenberg 25. For a uniform emitter with an effective aperture of up to 80°, the ratio of the length to width of the -6 dB zone is given by,

\[
\left( \frac{x}{z} \right)_{-6dB} = \frac{\alpha_m}{0.61(1 - \cos \alpha_m)} \quad (3.57)
\]

where \(\alpha_m\) is the angle of aperture. If the width of the EMAT -6 dB zone is assumed to be half as large as the distance between the first zeros of the uniform characteristic, figure 12, then,

\[
z_{-6dB} = \frac{1.83ct_{1/2}}{\alpha_m} \quad (3.58)
\]

where \(t_{1/2}\) is the half width of the acoustic pulse (figure 3) and \(c\) is the speed of sound. Neglecting reflection, damping and non linear effects, the theoretical pressure gain \(G_p\) is given by,

\[
G_p = \frac{2\pi}{3} \frac{F}{ct_{1/2}} \int_0^{\alpha_m} \phi(\alpha) \sin \alpha d\alpha \quad (3.59)
\]

where \(\phi(\alpha)\) is the distribution function for the pressure amplitude in the plane of the lens and satisfies the condition \(\phi(0) = 1\) and \(F\) the focal length of the lens. In the case of a polystyrene lens the reflection and damping losses account for a reduction in effective pressure at the lens of approximately 15%.
3.5 Pressure measurement

The measurement of acoustic fields in a great many cases can be facilitated by the use of a transducer. The name transducer is applied to devices which convert energy from one form to another, an example is the electromechanical transducer, which converts electrical energy to acoustic energy and vice versa. This particular type of transducer can be further grouped according to the principle of operation, which include electrodynamic, electrostatic, piezoelectric, ferroelectric and magnetostatic devices. In many cases a transducer may be capable of both transmission and reception, those used solely for underwater reception are called hydrophones. The design and selection of a transducer will depend primarily on the application, those used for the measurement of acoustic fields generated by shock wave sources such as the EMAT, require very specific properties. One of the major requirements of all hydrophones and that laid out in the AIUM/NEMA and IEC standards, is that the active element of the device is comparable to, or smaller than the wavelength. Whilst a specific property would almost certainly include the capability of withstanding the high pressures found in the focal regions of these devices. In addition, a stable and linear voltage response over a large dynamic range of pressures and a broad bandwidth with a smooth frequency response would also be necessary. Although in the latter case the high frequency harmonic content (>10MHz) of an EMAT shock wave is low and thus contributes very little in the way of energy to the shock wave. Even if the above requirements are met, the absolute value of pressure cannot be determined with certainty, because of (1) interference due to acoustic reflections from shock waves entering and exiting the transducer; (2) the interference with the propagation of the shock wave by the transducer package; (3) calibration uncertainty under such conditions and (4) the strong pressure dependence on the alignment of the transducer surfaces to the shock front.

One of the most common electroacoustic transducers in use today for the measurement of ultrasonic fields is the piezoelectric device. Their many advantages over other types of device include a compact and rugged design, high efficiency and a large operating temperature range. These devices typically use naturally occurring piezomaterials such as quartz and Rochelle salts, man made ceramics such as barium titanate and piezopolymers such as polyvinylidene fluoride (PVDF), vinylidene fluoride (VDF) and trifluoroethylene (TRFE). In recent years piezopolymers have been the subject of much research, there many additional advantages over piezoceramics have enabled novel transducer designs. Specific advantages of the polymer PVDF include an acoustic impedance similar to that of water and a frequency response from DC to
GHz. It also has a vast dynamic range from $10^{-5}$ to $10^8$ Nm$^{-2}$, a high dielectric strength (75 V/μm) and resists moisture, oxidants and most chemicals. In addition, it can be cut, formed and glued.

### 3.51 Piezoelectricity

Piezoelectricity describes the generation of electrical polarisation in a substance due to an applied mechanical stress or the deformation of that substance due to an applied electric field. The first of these two effects was initially discovered in 1880 by the brothers Curie and is known as the direct piezoelectric effect. The latter, discovered soon afterwards in 1881 is known as the inverse piezoelectric effect. In natural piezomaterials the effect is a property of the crystal structure and specifically the lack of symmetry of the molecules that form the crystals. In unstressed conditions the positive and negative charge centres of the crystal are coincident, figure 16a, thus there is no resultant dipole moment. However, when the crystal is compressed (figure 16b) or subjected to a tensile stress (figure 16c), charge separation occurs producing electrical dipoles within the crystal, leading to the production of surface charges.

![Figure 16 Piezoelectricity in quartz.](image)

If the stress is the result of a sound wave then an alternating voltage is produced which has the same frequency as the wave. Another group of materials that behave to all intents and purposes as if they were piezoelectric are ferroelectrics. Examples of this group of materials include PVDF, and certain titanates and zirconates, such as lead zirconate titanate (PZT). Normally ferroelectric materials are composed of a large number of randomly oriented polarised domains. If the material is heated to above its Curie temperature and then cooled in the presence of a strong electric field, the polarised domains can be made to line up with the applied field. This bulk polarisation of the material gives rise to a permanent electric dipole and thus strong piezoelectric properties. With ageing the remnant polarisation of the material slowly decreases, in time the material will return to its unpolarised state. For PVDF this loss of volume
polarisation has been found to decrease exponentially, with a time constant of the order of 100 years. A complete discussion of the manufacturing processes and polling procedure involved in the production of PVDF can be found in a study by Chandra.

In order to describe the various electrical and mechanical properties of piezomaterials, a number of coefficients are used. The piezoelectric stress constant $g_{ij}$ relates the electric field developed under open circuit conditions by the film, to the stress applied along a specific axis. In contrast, the piezoelectric strain constant $d_{ij}$ relates the change in the materials dimension along a specific axis, under stress free conditions, to the applied field. The electromechanical coupling factor $k_{ij}$ is a measure of the materials ability to convert electrical energy to mechanical energy and vice versa. Because piezomaterials are anisotropic their electrical and mechanical properties differ depending on the axis of the applied field or mechanical stress and strain. To identify the axis and a particular mode of operation, the two subscripts $i,j$ are used, the first, $i$ denotes the vector component of the electric field and the second $j$ the component of stress or strain. By convention the materials axis are identified by the numerals 1,2,3, where 1 corresponds to the length, 2, to the width and 3, to the thickness. The most common modes of operation are the thickness expander mode $(i,j) = (3,3)$ and the thickness shear mode $(i,j) = (3,1)$. From theory the induced open circuit voltage $V$ resulting from an applied compressive stress $X$ (Nm$^{-2}$) for a piezomaterial operating in thickness mode ($g_{33}$) is given by,

$$V = -g_{33}X_3 t$$

(3.60)

where the mechanical stress $X$ is positive for a tensile stress and negative for compressive stress and $t$ is the material thickness. For PVDF the thickness mode piezoelectric stress constant $g_{33} = -339 \times 10^{-3}$ Vm$^{-1}$/Nm$^{-2}$ and $X_3$ is negative (compressive stress). A comparison between the coefficients $g$ and $d$ for PVDF and other piezomaterials, show that PVDF is ideal for underwater reception but due to its low Q factor $\approx 1.4$ a relatively weak electromechanical transmitter. The fundamental half wavelength resonant frequency $f_0$ of a piezomaterial can be found from,

$$f_0 = \frac{v_s}{2t}$$

(3.61)

where $v_s$ is the speed of sound along a particular axis in the PVDF, in thickness mode $v_s = 2.2 \times 10^3$ ms$^{-1}$. 
3.52 Hydrophone design

The initial discussion on pressure measurement, specifically hydrophones suitable for acoustic shock fields, identified several ideal operational parameters. In practice the construction of such a hydrophone is difficult and thus optimisation through compromise and making trade offs between the particular parameters is often more appropriate. The first design objective is to match the hydrophones characteristics to the particular application, which in general can be grouped into either broadband or narrowband applications. For shock wave measurements a broadband device designed for optimal impulse response is the most suitable, since a relatively broadband frequency domain response and compact time domain response are required. A typical impulse response hydrophone therefore requires a low mechanical Q. One method of achieving this is to back the piezomaterial with a matched impedance high loss material. The lossy backing has the effect of spoiling the mechanical Q at resonance. However, impulse response is achieved at the expense of insertion loss and thus sensitivity. Since now, as well as reflection losses due to impedance mismatches at the hydrophone interface, a large fraction of the ultrasonic energy will be absorbed by the lossy backing. It is clear from the above, that the choice of piezomaterial and its mode of operation will be a very important factor in the design of a hydrophone. PVDF has already been identified as having an acoustic impedance near to that of water. At $3.9 \times 10^6$ kgm$^{-2}$s$^{-1}$ (thickness expander mode), its impedance is a sixth that of PZT ceramic. In addition its low mechanical Q, large piezoelectric stress constant and high sensitivity make it one of the most suitable materials for the construction of an impulse response hydrophone.

The high insertion loss characteristic of this type of hydrophone can be reduced by minimising the impedance mismatch at the water / hydrophone interface. To achieve this a $\lambda/4$ matching layer, which serves as an acoustic impedance transformer, can be placed between the water and the piezomaterial. Its $\lambda/4$ thickness ensures that any internal reflections are returned in phase, preventing the production of standing waves. Optimum impedance matching is obtained when:

$$Z_m = \sqrt{Z_p Z_w} \quad (3.62)$$

where $Z_m$, $Z_p$ and $Z_w$ are the acoustic impedances of the matching layer, piezomaterial and water respectively. The choice of matching material is generally a compromise between acoustic impedance, acoustic attenuation, its mechanical properties and its
sensitivity to the acoustic environment. An estimate for the mechanical $Q$ of a hydrophone at the half wavelength resonant frequency can be obtained from\textsuperscript{34},

$$Q_{\text{mech}} = \frac{\pi}{2} \left( \frac{Z_p}{Z_w + Z_b} \right)$$

(3.63)

where $Z_p$ and $Z_w$ are as equation 3.62 and $Z_b$ is the acoustic impedance of the backing material.

### 3.521 Prototype hydrophone designs

In the early stages of the research, transducer measurements were used to assist in the design and selection of suitable diaphragm materials and lens formats. Later on they also played a major role in the characterisation of the acoustic field. This essential role meant that the transducer design had to be continually updated to take account of the changing requirements. The first prototype hydrophone, shown figure 17a, was inherited from an earlier project\textsuperscript{35}.

![Figure 17 The hydrophone prototypes.](image)

Its design is basically that of a brass conical horn coupled to a resonant cylinder. The horn and the resonant cylinder have the effect of amplifying the incident acoustic wave. The active element of the hydrophone was a thin disk of PZT (EC-64 type I) 4 mm in diameter, which was located within a hollowed out part of the resonant cylinder. Electrically, one side of the PZT was directly coupled to the body of the hydrophone, the other, via a partly insulated brass bar that formed the backing, to the input of the oscilloscope. The design only allowed the hydrophone to be partially submerged and was thus used vertically with only the horn in the water. This restricted its use to only the very early stages of the EMAT development, when the shock waves were emitted into a vertical column of water prior to the adoption of the horizontal tank format. A typical voltage trace produced by the transducer for a focused shock
wave is shown in figure 18. The trace shows that a very respectable voltage in excess of 40 V pk-pk was produced, despite the transducers considerable reflection loss of around 85%.

Figure 18 The voltage trace produced by the horn transducer, (1 horizontal div=10 μs, 1 vertical div=5 v).

The complex waveform is dominated by the effects of resonance and multiple out of phase reflections, as a result the rather simple form of the shock wave is lost. However, by considering just the first cycle of the waveform an approximate value for the maximum amplitude of the wave could be made.

The first fully submersible purpose built impulse response hydrophone was the PVDF-PTFE hydrophone, shown figure 17b. This was the first of the prototypes to use the piezopolymer PVDF. In this design a 5 mm diameter, 110 μm thick sample of the film with a thickness mode resonance of 10 MHz, was coupled to a similarly sized 30 mm long rod of polytetraflouroethyene (PTFE). PTFE was chosen because of its similar acoustic impedance \( Z=3\times10^6 \text{ kgm}^{-2}\text{s}^{-1} \) and its relatively lossy characteristics. The length of the PTFE rod ensures that no longitudinal or shear waves propagating in the backing are returned during the first oscillation cycle of the shock wave. The electrical connections to the PVDF film were made by way of silver loaded epoxy paint, which was applied together with the connecting wires to the elastomer based silver electrodes. When dry, the wires and the electrodes were bound along the length of the rod with cotton thread prior to being permanently fixed in place with epoxy resin. Acoustic coupling gel was then applied to the surface of the PVDF before it was inserted into the hydrophone body and sealed. A major disadvantage of this type of device is of course its insertion loss. To reduce this to a minimum in this first device, a polystyrene end window with an acoustic impedance \( Z \) of \( 2.5\times10^6 \text{ kg m}^{-2}\text{s}^{-1} \) (table 1)
was incorporated into the polystyrene body of the hydrophone. This provided an optimum impedance match between the water and PVDF and thus reduced the reflection loss of the hydrophone to less than 10% of the incident energy. The voltage trace produced by the hydrophone for a focused shock wave at a point on the axis of the EMAT is shown in figure 19.

![Figure 19 The voltage trace of a shock wave using the PVDF-PTFE hydrophone, (1 horizontal div=2μs, 1 vertical div=5 v).](image)

The transducer has clearly reproduced the form and relative amplitude of the various components of the shock wave, although it has not been able to resolve its leading edge. This was thought due to the effect of the quarter wave end window, which by its very design limited the operational bandwidth of the hydrophone. The loss of sensitivity over its often short operation lifetime was also found to be a problem. Similar effects have been reported widely in similar devices using PVDF. The effect in some cases has been attributed to the action of cavitation on the surface of the PVDF. In others, it is thought due to delamination of the elastomer electrodes as a result of the strong forces exerted on the PVDF by the shock wave. Although no study of the damage on the PVDF was undertaken, it is thought likely that the latter case was the cause of the downward drift in sensitivity in our device.

In an attempt to faithfully reproduce the shockfront and improve the operational lifetime, a third hydrophone prototype was constructed. The PVDF-PVDF hydrophone shown figure 17c, was based on a 110 μm thick sample of the PVDF piezofilm with a 3 mm active diameter. Directly coupled to this was a matched impedance backing of unpolarised PVDF. The near perfect impedance match had the effect of further broadening the operational bandwidth of this device compared to the previous designs. The electrical connections to the piezofilm electrodes were made in the same way as that described earlier, since satisfactory results were always obtained despite it being
fiddly and time consuming. In an attempt to increase the dynamic strength and thus the
longevity of the hydrophone, the piezofilm and the end of the rod were encapsulated in
epoxy resin. This not only reduced the effects of delamination but also acted as an
acoustic impedance transformer, reducing the insertion loss. The reflection loss for
this device was calculated to be less than 13% of the incident energy. On sealing the
device into the housing the epoxy resin also formed part of a thin end window, which
had to be polished flat and parallel to the piezofilm surface. In the preparation of the
epoxy care had to be taken to prevent the formation of bubbles, since the end window
should ideally be homogenous and free from any imperfections. The predicted
directional sensitivity of the hydrophone, which can be assumed equivalent to that of a
flat piston receiver, for an acoustic frequency of approximately 200 kHz will be
independent of angle. The hydrophone output for a fully developed shockfront is
shown in figure 20.

![Image of hydrophone output](image)

Figure 20 The voltage trace of a shock wave using the PVDF-PVDF hydrophone, (1
horizontal div=2 μs 1 vertical div=5 v).

The trace clearly shows the very fast near vertical slope of the shockfront followed by
a rarefaction phase and a further compression cycle. At various points along the
waveform minor peaks are seen, these are thought due to mode conversion within the
lens (section 3.4) or from internal reflections within the hydrophone. The measured
rise time, although not resolved in this trace is 300 ns. Other studies of similar EMAT
shock fields using superior PVDF membrane hydrophones have measured rise times
of 120-30 ns. However, the maximum rise time capable of being measured is
ultimately limited by the piezofilm resonance, which for 110 μm thick piezofilm is
about 100 ns. This much improved response and the hydrophones ability to withstand
extended periods in the focal region of the system proved to be a firm basis on which
all further results could be taken.
3.53 Hydrophone calibration

To accurately determine the temporal and spatial pressure distribution of the acoustic field generated by the EMAT, the hydrophone sensitivity must first be calculated or measured. The sensitivity is defined as the ratio of the open circuit voltage at the terminals to the sound pressure and can be expressed in units of \( \text{V Pa}^{-1} \) or in decibels relative to \( 1 \text{ V \mu Pa}^{-1} \). For recording purposes the PVDF-PVDF hydrophone was connected via its 0.7 m coaxial cable to the input of a Tektronix 466 analogue storage oscilloscope, which has an input resistance of \( 1 \text{M\Omega} \) and a capacitance of 20 pf. The simplified voltage mode equivalent circuit of the piezofilm, cable and oscilloscope are shown in figure 21.

![Figure 21 The equivalent circuit for the piezofilm and the input impedance of the oscilloscope.](image)

The piezofilm is represented by the open circuit voltage source \( V \) and the series capacitance \( C_r \). In parallel to this is the oscilloscope input resistance \( R_i \) and the lumped capacitance \( C_T \). By inspection the terminal voltage \( V_o \) and thus the theoretical sensitivity is given by,

\[
V_o = \frac{C_f}{C_f + C_T} V \tag{3.64}
\]

where the induced open circuit voltage \( V \) for an applied stress \( X \) is given by equation 3.60. Thus the calculated sensitivity for a 110 \( \mu \text{m} \) thick sample of PVDF, operating in thickness mode \((g_{33})\), with an applied stress of \( 1 \text{ Pa} \), is \( 37.3 \ \mu \text{VPa}^{-1} \). Given the hydrophone reflection loss of 13\% and an approximate value for the film capacitance \( C_f \) of 6.81 pf and also that \( C_T = C_i + C_c \), where \( C_i \) (the oscilloscope input capacitance) = 20 pf and \( C_c \) (the capacitance of the coaxial cable and BNC connectors) = 70 pf, then the calculated hydrophone sensitivity is \( 3.30 \ \mu \text{V Pa}^{-1} \) or \(-229.6 \text{ dB}\). The estimated value for the mechanical \( Q \) of the hydrophone is found from equation 3.63 to be approximately 1.14. As a comparison the calculated value was compared with the
measured value of sensitivity obtained using the transducer test facility at the GEC Marconi research centre Chelmsford. Figure 22 shows the measured sensitivity for a range of frequencies between 0.5 MHz - 10 MHz for the PVDF-PVDF hydrophone.

![Graph showing PVDF-PVDF hydrophone sensitivity.](https://via.placeholder.com/150)

Figure 22 PVDF-PVDF hydrophone sensitivity.

The calibration plot shows a marked fall off in the sensitivity of the hydrophone from 1 MHz - 10 MHz. Even allowing for the increase in the amplitude reflection coefficient$^{37}$ of the film with frequency, a small overall increase in sensitivity might have been expected due to the effect of the film resonance. The reason for the rapid fall off only becomes clear when the dimensions of the transducer source and thus its focal region are considered. For these calibration tests a transducer with an aperture of 10 mm and a focal length of 50 mm was used. Analysis of the width of the focal zone using the Rayleigh criterion, indicates that above 1 MHz the difference between the 3 dB width of the focal zone and the 3 mm diameter active element of the hydrophone is likely to lead to significant spatial averaging of the pressure. Further evidence of the unsuitability of this hydrophone for measuring frequencies above 0.5 MHz is seen in the IEC standards$^{27}$, this is especially so where there is a significant harmonic content due to finite amplitude distortion$^{38}$. Correction factors for the effects of non-linear distortion and spatial averaging can be found in a publication by Preston et al$^{39}$. Because insufficient calibration data exists for the hydrophone the spatial averaging correction factor above 1 MHz could not be calculated with any certainty, but it is
likely to be significant. The comparison between the calculated sensitivity and the measured value at 0.5 MHz show they differ by approximately a factor of 2. As far as the calculated result is concerned, the high value may be due to an underestimation of the stray capacitance or other electrical influences on the circuit, attenuation is also likely to effect the sensitivity. By contrast the measured value may be low simply because of the factors described. The measured sensitivity value may have been greater and more accurate if a reading could have been taken at 200 kHz, since this is the main frequency component in the EMAT pulse. The sensitivity of the hydrophone might therefore lie somewhere between the upper and lower limits of the calculated and measured sensitivity respectively. However, to be on the safe side an absolute value of sensitivity for the hydrophone of 1.2 \( \mu \text{VPa}^{-1} \) was adopted for all measurements.

3.6 The electromagnetic acoustic transducer II design

3.61 The EMAT coil

Earlier in this chapter we outlined the principles behind the operation of the EMAT and also suggested several ways in which the conversion of electrical energy to kinetic energy in the diaphragm could be maximised. From the discussion it is clear that the coils characteristics will also play a major role in this, i.e. increasing the radiated pressure is not as simple as just increasing the operating voltage and thus the discharge current, although this does work very well. In addition to the resistivity of the diaphragm and the coil/diaphragm separation, the diameter of the wire, number of turns and the diameter of the coil are also important, this fact is apparent in equation 3.35. To see what effect changing these factors have, it is helpful to understand how they effect the coil and the EMAT as a whole. This has been discussed and analysed in some detail by Alcock\(^\text{12}\), although the accuracy of his results cannot be confirmed. One of the ways in which changing the above parameters can effect the characteristics of the coil is through self induction. Here, the magnetic flux generated around the conductors of the coil by the flow of current, induces an e.m.f in that same coil which from Lenz's law opposes the changes in current that caused it. The induced e.m.f, which is proportional to the rate of change of current is given by,

\[
V_i = -L \frac{dl}{dt} \tag{3.65}
\]

where the constant of proportionality L is the self inductance or simply the inductance of the coil. Starting with the Biot-Savart law\(^\text{40}\) it can be shown that this inductance is
proportional to the square of the number of turns and the linear dimensions of the coil, both a measure of conductor length. The magnetic flux also gives rise to induced e.m.f's in the diaphragm through the effects of mutual induction. These are similarly proportional to the rate of change of current in the coil and related by,

\[ \nu_2 = m_{12} \frac{dl}{dt} \]  

(3.66)

where \( m_{12} \) is the mutual inductance of the diaphragm due to the coil. As before it can be shown that this mutual inductance is proportional to the number of turns in both the coil and diaphragm (\( N=1 \)) and the separation between them. With reference to section 3.3, this suggests that the self inductance of the coil should be minimised whilst the mutual inductance is maximised. Hence, for the greatest efficiency and thus pressure we require a coil with a large number of turns, a small radius and the smallest possible coil diaphragm separation. The limiting factor on the radius of the coil is of course the reduction in lens focusing efficiency for increasing values of \( \lambda/D \). With this in mind the final coil had 20 turns and a radius of 40 mm. The close proximity of the windings required the coil to be wound with high purity copper conductor (1.5 mm \( \Omega \)), and insulated with a tough (class H) two-layer enamel (polyester/polyamide-imide) to prevent electrical breakdown.

Initially all the EMAT coils were wound laboriously by hand therefore, only in a very few cases were the results satisfactory, this was especially so for coils of greater than 10 turns (\( N=10 \)). For the larger diameter coils (\( N=20 \)) hand winding usually led to highly asymmetric spirals in which the inter turn winding separation varied across the coil. Using such coils would have almost certainly resulted in the non symmetric excitation of the diaphragm and thus produced highly undesirable far field pressure distributions. To rectify this problem and at the same time speed up the coil winding process, a lathe mounted coil jig was developed. The jig, which was rotated slowly by the lathe, consisted of two large diameter plastic plates held a distance equal to the exact diameter of the winding wire apart. A single hole near the centre of the jig between the two plates, provided the winding anchor point for the wire at the centre of the spiral.

To wind a coil, a sufficient length of the winding wire was first threaded through the anchor point, this was to be used later to make the high tension (HT) connection to the coil. The jig was then slowly rotated and the wire fed under tension until a spiral of the required number of turns had been wound. The finished coil was then bonded using epoxy resin to the shot blasted surface of the coil support or acoustic backing. The
backing, an 80 mm diameter disc of 16 mm thick float glass presented an acoustically hard electrically insulating surface on to which the coil could be mounted. In the centre of the disc a 4 mm hole allowed the HT connection to pass through to the rear of the mounting flange (see figure 5), where both ends of the coil were terminated. When operating, the EMAT coil is subjected to severe electrical and mechanical stress. The mechanical stress is due to the force experienced between any two current carrying conductors in a magnetic field. In the spiral coil this force is essentially attractive (right hand screw rule and Flemings left hand rule) and proportional to the magnitude of the current and the reciprocal of the distance between the two conductors. In simple terms the coil will experience severe and very rapid contractions during the period in which the discharge current flows. Over time, this could lead to the disintegration of the coil. To lessen the effect and prevent the ingress of moisture, the coil was hermetically sealed in a flat but thin slab of epoxy resin. The coil was also protected from being short circuited by placing an insulating acetate sheet between the coil and metallic diaphragm. The EMAT with the diaphragm and insulating sheet removed to expose the spiral coil is shown in figure 23.

Figure 23 The EMAT with its spiral slab coil exposed.

3.62 The Diaphragm

The selection of the correct diaphragm material was a particularly important design consideration, since the choice not only effects the electroacoustic efficiency of the EMAT but also its operational reliability. In an earlier discussion it was suggested that the resistivity of the material should be as low as possible, inorder to maximise the force on the diaphragm. Further requirements are that the material should be flexible and able to withstand that force without being permanently deformed. To get a feeling
for the degree to which stress might effect the diaphragm, it was helpful to consider how far the diaphragm is displaced during operation. By considering the relation between the pressure $P$ and the bulk modulus $K$, (Equ.3.01) we can estimate the change in thickness $\Delta s$ of a disk of water of thickness $s$, in the vicinity of the diaphragm required to produce a particular pressure $P_m$. Hence from equation 3.01 we can write,

$$\Delta s = \frac{P_m s}{K}$$  \hspace{1cm} (3.67)

where the bulk modulus $K$ for water is 2 GPa, $P_m$ is the peak plane wave pressure $\approx$ 5 MPa and $s$ is taken to be equal to the wavelength of the acoustic pulse $\approx 9 \times 10^{-3}$ m. From this we can see that $\Delta s = 22 \times 10^{-6}$ m is actually very small, indicating that the diaphragm displacement is also very small. Even with such small displacements, it was clear from observations that the elastic limit of several sample diaphragms had been exceeded, resulting in plastic deformation. Repeated deformation often leads to fatigue, in which the internal molecular structure of the material changes, leading to a reduction in the material ductility. This was seen in several cases and ultimately lead to the materials failure. The ideal diaphragm material must therefore posses an extremely low bulk resistivity, coupled with sufficient yield strength and ductility to withstand the fatigue of repeated operation. Materials that have the lowest bulk resistivity at room temperature such as silver and copper also have relatively low yield strengths. These depend on the temper state of the material but are in the region of 100-300 MPa. Other commercially available metals, such as brass and aluminium, have better yield strengths but bulk resistivities that are many times higher.

To gain a better understanding of what these figures for resistivity and yield strength meant in the operational environment, six samples were selected as prospective diaphragm materials. Diaphragms of copper, copper-polyester composite and brass were subjected to 500 shots or to destruction, whichever occurred first. In all cases the far field pressure distribution was measured to determine which diaphragm produced the highest pressure and thus had the highest electroacoustic efficiency. Figure 24 shows the pressure amplitude with radial distance for a range of test materials. The results indicate, as expected, that the pure copper samples with their lower resistivity yield the highest conversion efficiencies and thus the highest pressures. Of the samples tested the 0.05 mm thick copper foil arguably has the highest overall pressure distribution, although in its annealed form (figure 25 bottom) it often suffered catastrophic failures.
Figure 24 The far field pressure distribution profiles for different diaphragm materials.

The long term survivability was only marginally improved when the material was used in its half hard and fully tempered states. In the half hard state, substantial rippling of the surface occurred (figure 25 top left), whilst in its fully tempered form the spiral pattern of the coil windings could be clearly seen (figure 25 top right). In both cases the damage resulted in the reduction of the far field pressure amplitude as the magnetic coupling between the coil and the diaphragm became less efficient. The copper polyester composite samples were produced from material commonly used in electronics for the production of flexible bus connectors. Its performance was in general quite good, although in the thinner samples there was a tendency for delamination to occur, sometimes within only 30 shots (figure 25 top centre). The delamination damage was similar to the damage observed in the annealed copper diaphragms. In both cases this was often accompanied by the formation of an arc discharge across the damaged region. The brass and the 0.075 mm thick fully tempered copper foil suffered none of the fatigue problems seen with many of the other diaphragm materials and thus produced constant outputs throughout their test periods. Since both samples clearly offered a similar level of operational reliability, the final choice favoured the sample with the highest efficiency. This was of course the 0.075 mm thick copper diaphragm. To date a diaphragm of this type has produced in excess of 5000 shock waves over an operational period of 24 months.
3.63 Lens design and manufacture

The intensity that can be generated from the surface of the EMAT will ultimately be limited by its design and its operational environment. The poor inherent electroacoustic efficiency of these devices means that the required intensity can only be achieved by a substantial increase in operating voltage. This would almost certainly necessitate improvements in the mechanical strength and insulation resistance of the EMAT coil, not to mention a stronger diaphragm, if the earlier damage is to be avoided. Cavitation would also become a problem, consuming a large part of the radiated energy and systematically damaging the radiating surface. The answer to this problem is of course the use of an acoustic lens or similar focusing system. Using such a lens it is possible to generate high intensities within a well defined region or area of interest. The material requirements and design criterion for acoustic lenses have already been discussed earlier in this chapter. When considering these requirements it can be seen that plastics such as perspex and particularly polystyrene are ideally suited to this application. The fact that polystyrene is easily available, has an acoustic impedance nearer to that of water than perspex and a lower acoustic absorption, (the predominant loss mechanism) meant that it was used in preference to perspex.

Chemically, polystyrene is produced by adding a peroxide catalyst (Azo bisisobutryo nitrile) to styrene monomer in an oxygen free nitrogen atmosphere, at temperatures between 100 - 200 °C. During the reaction 95 - 100% of the monomer will be
converted to the polymer, any excess monomer at this stage is removed in a partial vacuum and recycled. The polymer melt is then extruded as filaments and cooled in a water bath, prior to being chopped and packaged. Polystyrene is completely amorphous and therefore has excellent optical clarity. Its bulky ridged polymer chains also make it quite ridged but at the same time brittle. On the downside, polystyrene is attacked by organic solvents and has poor UV and oxygen resistance.

Having chosen the lens material, potential manufacturing methods then had to be considered. Amongst those considered as viable in what was a very limited manufacturing environment were, (1) the production of the lens using the styrene monomer and catalyst; (2) injection moulding; (3) machining the profile from bulk polystyrene and (4) heating the chopped or bulk polystyrene in a lens mould using an industrial oven. The first method was judged far too complex due to the strict temperature and environment requirements. The second method was similarly rejected on the grounds of the potential expense of a two part mould and the unavailability on site of a machine capable of delivering a shot weight of approximately 500g. The availability of a limited supply of bulk crystal grade 303 polystyrene from the Huntsman chemical company meant that the machining route was actively considered. However, stresses present within the bulk material resulted in the disintegration of the partly machined lenses in almost all cases and eventually this method was also rejected. The fourth option in which chopped or bulk polystyrene was melted in a lens mould, was also far from satisfactory. The polystyrene's high viscosity, even at temperatures above 200 °C, meant that for satisfactory results extended periods in the oven were required. This was particularly so when chopped granules of polystyrene were used, since air pockets trapped by the melting granules often took an additional 3-4 hours to rise the surface of the viscous polystyrene. With such large periods of time required to produce the desired homogeneity, large scale decomposition of the polystyrene often occurred, this was particularly apparent at the surface due to its contact with the air.

To a certain extent these problems were reduced in the later lenses by the use of bulk polystyrene. Despite it taking slightly longer to melt the block, the overall time in the oven was reduced due to the absence of air bubbles in the material. Further improvements in manufacturing time and lens quality were obtained by heating the polystyrene in a rarefied atmosphere of inert argon. This allowed higher temperatures typically 250 °C to be used without the danger of material decomposition.
3.631 The acoustic lenses

The first prototype acoustic lens was of a relatively simple plano-concave design, similar to that shown in figure 26a. The choice of a single spherical concave surface for a 100 mm lens of this type, would have almost certainly resulted in a large spherical aberration. A better solution was therefore a spherical approximation of an elliptical form, in which two merging spherical radii approximated part of the profile of an ellipse. The advantage being that elliptical profiles are not effected by spherical aberrations to the same extent as a conventional spherical lens surface. The final aspheric surface design, which was modelled on an ellipse with major and minor axis of 168 mm and 130 mm respectively, had a centre radius of 50.5 mm followed by a transition to 56.5 mm at a height of 32.5 mm above the optical axis of the lens. The theoretical focal length for the 10 mm thick lens was calculated to be approximately 136 mm, giving the lens an effective aperture angle $\alpha_m$ of 21°, where $\alpha_m$ is the maximum angle made by a marginal ray with the optical axis. The theoretical pressure gain $G_p$ from equation 3.59 for the lens (neglecting reflection and absorption losses), given that $t_{1/2}$ for the unfocused shock wave is 2 $\mu$s and the value for the integral is 0.0522 using the pressure distribution shown in figure 24 (0.075 mm copper), was calculated to be approximately 5. The estimated 6 dB width of the focal zone (Equ.3.58) was 14.75 mm.

![Figure 26a, The plano-concave lens design and b, the bi-concave design.](image)

To manufacture the lens, the negative form of the required surface profile was machined onto an aluminium template. This then formed part of a simple self centering mould into which the required quantity of polystyrene granules were then placed. When the polystyrene had completely melted and was totally free of bubbles it was removed from the oven and allowed to cool. The lens blank was then removed from the mould and the planar surface of the lens and the various fixing hole machined on. The second lens designed and manufactured for use with the EMAT was of a much improved bi-concave design, as shown in figure 26b. The two spherical surfaces were
chosen using the 'best form' approach, (see section 3.4.11) in which spherical aberration is reduced by the correct choice of surface radii. Thus for a 100 mm focal length lens with an object distance set at infinity (planar waves), the required radii for the first surface \((r_1)\) and the second surface \((r_2)\) were -136 mm and 51 mm respectively. The associated sign convention indicates that both surfaces are concave as expected. The effective aperture angle \(\alpha_m\) for the 120 mm diameter lens was calculated to be 36° and the theoretical 6 dB width and pressure gain (given that value of the integral of Equ.3.59 is 0.078 for the distribution of fig.24), to be 8.64 mm and 5.5 respectively. Ray tracing calculations for a number of representative rays, indicated that the longitudinal spherical aberration (LSA) would be in the region of 20 mm for this lens. However, its improved focusing ability was partially offset by increases in mode conversion, reflection and absorption losses within the lens. These were calculated to be approximately -1.7 dB at the centre of the lens and reach a maximum of -13 dB for marginal rays at the periphery. The relative complexity of the bi-concave lens meant that it was necessary to produce it in two parts. Each part, which was effectively a plano-concave lens, was manufactured in an inert argon atmosphere from bulk polystyrene. Apart from this, the manufacturing process was the same as described earlier and used the original self centering mould and the appropriate aluminium template. The two planar surfaces were then machined and polished flat prior to being joined with commercially available polystyrene cement.

3.64 Operational characteristics

Possibly the most interesting part of any design project is seeing how the actual operational characteristic of the device compare with the theoretical predictions. For the EMAT this meant investigating not only its electrical characteristics but the properties of the focused and unfocused acoustic fields, whilst simultaneously testing the hydrophone. Initially then, we shall look at the electrical characteristics of the EMAT under discharge conditions and estimate the radiated acoustic pressure. From theory (Equ.3.30) the discharge current in the circuit can be seen to be dependent upon the operating voltage and the circuits impedance. The maximum operating voltage is limited by the voltage rating of the 0.544 \(\mu\)F discharge capacitor at around 18 kV, thus rating the discharge capacity of the discharge system (Equ.3.26) at 88 Joules. Figure 27 shows the discharge voltage measured across the EMAT during operation. The highly damped voltage transient has a period of 6 \(\mu\)s (166 kHz), which is in good agreement with the predicted relation of figure 10.
To determine the discharge current it was first necessary to calculate the complex impedance of the coil, which is essentially inductive. Thus rearranging equation 3.31 for L, where the measured value of $R = 0.036\Omega$ and the angular frequency $\omega$ from figure 27 is $1.047 \times 10^6 \text{ rad s}^{-1}$, we obtain $1.67 \mu\text{H}$, and therefore a coil impedance of $0.036 + j1.75$. From ohms law, the highly damped current ($\delta = 0.25$) has a maximum value of around 9.1 kA, followed by a negative value of 5.15 kA and a second peak of 2.28 kA. The peak radiated pressure $P_m$ (Equ.3.35) in the vicinity of the EMAT diaphragm is therefore calculated to be approximately 13 MPa. However, this figure represents the best case situation since the expression takes no account of the diaphragm efficiency, which has been shown to vary depending on the choice of material. Therefore, equation 3.35 can only be considered as no more than a rough approximation, the actual pressures are considerably lower. Pressure measurements made at the same position along the axis of the EMAT for a planar and a focused acoustic wave are shown along with their corresponding frequency spectra in figure 28 a,b.

Figure 27 The discharge voltage across the EMAT, (1 hor div=5\mu s 1 vert div=5000 v).

Figure 28 The shock wave and frequency spectra for a, the planar wave (1 hor div=20 \mu s) and b, the expanded focused wave (1 hor div=2\mu s). In the frequency domain 1 horizontal div=0.1MHz.
The inverted appearance of the two shock waves (top of figure) is due to an error made when connecting the electrodes to the PVDF for this particular hydrophone. In this case, the shock front and the rest of the compression phase of the wave front is represented by a negative voltage and the rarefaction phase by a positive voltage. The temporal form of the wave front is not affected. Analysis of the two wave forms show that the focused wave (fig.2b) has a much faster rise time than that of the planar wave, as is expected from theory c.f. Equ. 3.08 & 3.09. The wave form distortion associated with the production of a shock front can also be seen in the frequency domain as a shift of energy from the low to higher harmonics. For the planar wave the dominant frequencies are below 100 kHz, whilst for the more distorted focused wave a shift above 100 kHz is apparent. This shift in the frequency spectra will also be apparent for a focused wave as the effects of non-linear distortion change the wave shape during propagation. Hydrophone measurements of a focused shock wave taken along the z axis at distances of 80, 110, 130 and 180 mm from the radiating surface of the EMAT (0, 30, 50 and 100 mm in front of the lens) are shown in figure 29abcd.

Figure 29 Shock wave development for a focused emitter, a, initial sinusoidal wave; b, developing shock front; c, fully developed shock front; d, ageing shock front (1 horizontal div=2 μs 1 vertical div=4.52 MPa).
Chapter 3: The generation and focusing of acoustic shock waves

The main feature of all the images is the shockfront, which is seen to steepen as the wavefront approaches the focal region and then age or become less steep as it moves away from the focal zone. After the initial compression phase, the pressure falls from its peak value to zero in around 3 μs and then goes negative achieving its peak negative value in about 4 μs after the shock front. This is followed by a second compression approximately 6 - 8 μs later. The negative or rarefaction phase of the shock wave is associated with diffraction and the production of edge waves originating from the periphery of the EMAT diaphragm. The duration of the negative phase of the wavefront is generally greater than the initial compression and is often seen to increase with increasing pressure. The peak positive and peak negative pressure for the bi-concave lens was measured and found to be approximately 19 MPa and 5 MPa respectively, giving a $P_+/P_-$ ratio of 4. The waveforms obtained at symmetrical positions on the z axis, either side of the focus and at similar positions on the y axis are identical i.e. the wave can be considered symmetrical about the axis of the EMAT. Pressures away from the EMAT axis are found to fall, and rise times increase with increasing distance from the axis. This occurs for both planar and focused wave forms although the difference is more marked in the focused case. A comparison between the force on the diaphragm (figure 10), and the temporal measurements of the planar and focused wave fronts, on the whole indicate very little similarity, except for the shock component itself. It is suggested that diffraction, mode conversion and transducer limitations may mask the second weaker diaphragm displacement that peaks 4 μs after the first.

Figure 30 a,b show the axial and radial pressure distributions for the planar and focused acoustic fields described earlier. A comparison between the calculated focal lengths for the two lenses and their axial pressure distributions fig.30a, suggest a relatively good correspondence. Although, there is some confusion over the actual position of the focal points, due to the relative uncertainty about the actual positions of the secondary principle planes of the two lenses. This is particularly so in the case of the large bi-concave acoustic lens. The length of the 6 dB regions for the bi-concave and plano-concave lenses are estimated to be 95 mm and 152 mm respectively. Similar measurements made on the radial pressure distributions of fig.30b, yield 6 dB widths of 8.38 mm and 17.41 mm for the two lenses. A direct comparison of these figures with those calculated using equation 3.58 in section 3.6, indicate an excellent level of agreement for the bi-concave lens (8.64 mm) and an overestimation of around 18% for the plano-concave lens (14.75 mm). However, since the principle assumptions stated for equations 3.58 & 3.59, namely harmonic oscillation and linear propagation are not satisfied, errors in the calculated results for focal geometry and pressure gain...
can be expected. The length to width ratio of the 6 dB zones for the two lenses are therefore estimated to be 11.33 and 8.73 for the bi-concave and plano-concave lenses respectively, and are in good agreement with studies of similar EMAT lens combinations\textsuperscript{42}.

An estimate of the focal gain for each lens can be made by considering the peak focal pressure and the peak plane wave pressure. The ratio of the two yield a pressure gain for the bi-concave lens of approximately 4 and 2.8 for the plano-concave lens. In contrast, the theoretical predictions indicate values of 5.5 and 5 respectively. The higher gain obtained for the bi-concave lens is consistent with its larger aperture and reduced longitudinal spherical aberration. Overall the similarities between the calculated and measured properties of the acoustic fields for the bi-concave lens, are indicative of the more rigorous approach applied to the design of this particular lens. Figure 31a-c show 2-D contour representations of the data presented in figure 30 for the planar and focused acoustic fields. The peak pressure measurement were made at 5 mm intervals radially and at 10 mm intervals along the beam axis. The maps clearly show that the focal regions are characterised by a wider pressure distribution in the direction of propagation than in the transverse direction.
Figure 31a The planar acoustic field.

Figure 31b The focused acoustic field for the plano-concave lens.

Figure 31c The focused acoustic field for the bi-concave lens.
The increase in shock wave rise time observed with increasing distance from the EMAT axis is shown for both planar and focused wavefronts in figure 32 a,b. The measurements were taken in the focal plane for the focused shock wave and at the same position on the axis with the lens removed for the planar wavefront. As the pressure falls towards the periphery of the EMAT the rise time can be seen to increase. The basis for this can be found by analysing equations 3.08 & 3.09, which show that with all other parameters being constant, a higher acoustic pressure will lead to a reduction in the shock formation distance $S$. Hence a wavefront with a non uniform pressure distribution may also have a non uniform shock formation distance, and thus regions of the wave in which the shock front is in a more advanced state of development. A comparison between the rise time for a particular pressure on the planar wavefront and the same pressure irrespective of position on the focused wavefront, indicates that at comparable pressures rise times for the planar wave are faster. A possible explanation for this may be attributed to the lens and the effects of non linear absorption, which tend to be greater for the lens and the focused wavefront than for the unimpeded planar wave. However, once the focused wavefront emerges from the lens rapid development of the shockfront occurs, especially along the axis where the pressures are the greatest. The theoretical shock formation distance $S$ for the planar wave can be estimated from equation 3.08, hence for a peak plane wave pressure of 4.66 MPa, $t_{1/2} = 2$μs and $B/A = 5$, $S = 189\times10^{-3}$ m. This distance is reduced in the case of the focused wavefronts (Equ.3.09) to approximately $85\times10^{-3}$ m and $102\times10^{-3}$ m for the bi-concave and plano-concave lenses respectively. In both cases it can be seen that a fully developed shockfront is produced prior to the wave reaching the focal zone of the EMAT lens combination.

One of the most important parameters of the shock wave is the total acoustic energy or energy per pulse. This can be estimated (Equ. 3.38) for a shock wave with a pressure $P(t)$ described by equation 3.36. Thus for a shock wave with a peak pressure ($P_m$) of 19 MPa and a period of 6 μs, the calculated peak energy density is approximately $106.5 \text{ Jm}^{-2}$. The total energy in an area equivalent to the 6 dB width of the bi-concave lens is thus approximately 8.5 mJ, giving an overall electroacoustic efficiency of 0.01%.

Because the EMAT has a finite radiating area it will have some degree of directionality (chapter 3.32). The angular divergence of the acoustic beam for the EMAT can be determined from its directivity function shown in figure 12. The actual angle is determined by the point on the function at which the pressure drops to 70% (-3dB) of its axial value. Thus from the characteristic $\beta R_\theta \sin\theta = 2.1$, the angle of divergence is therefore 5.75° and the beamwidth ($2\theta$) of the acoustic field is 11.5°.
The results obtained during the research have shown the EMAT to be a reliable and consistent source of shock waves. The shot to shot reproducibility of the device was found to vary by less than 5% of the peak pressure in both the planar and focused wavefronts. The main cause of this small pressure fluctuation is the variable characteristics of the arc discharge between the electrodes of the spark gap switch.

![Graphs showing radial pressure distribution and rise time for planar and focused waves.](image)

Figure 32 The radial pressure distribution and rise time for a, the planar wave and b, a focused wave.

However, it was found that regular maintenance and cleaning of the discharge electrodes was all that was required to maintain the optimum operating efficiency of the switch. The long term stability of the EMAT was not effected and no downward drift in output pressure has been observed. The EMAT's operational life time is unknown, however, to date after 24 months of operation no discernible damage has occurred to either the coil or diaphragm. Although, the biconcave lens has suffered the effects of water ingress, resulting in the partial separation of the two lens components. Hydrophone measurements of the acoustic field, particularly in the focal region have been found to cause a rapid downward drift in hydrophone sensitivity. This occurs even with the latest hydrophone design therefore, measurements of this region were kept to a minimum. Later measurements have been based on a schlieren technique, the theory of which shall be discussed in the following chapter.
Chapter 3: The generation and focusing of acoustic shock waves

References

27 Measurement and characterisation of ultrasound fields using hydrophones in the frequency range 0.5 to 15 MHz, IEC Draft standard 87 (secretariat)2 1987.


Shotton, K.C. Bacon, D.R. Quilliam, R.M. A PVDF membrane hydrophone for operation in the range 0.5 MHz to 15 MHz. Ultrasonics. 18:123-126, 1980.


CHAPTER 4

SCHLIEREN METHODS

4.1 Introduction

In 1856 Foucault proposed an excellent method for detecting the presence of inhomogeneous regions in optical components. The inhomogeneities within the glass components were often in the form of streaks or schliere, and it was after these that the technique derived its name. The schlieren method was later applied to acoustics by Toepler, who used it to observe shock waves generated by a spark discharge. Schardin then put the method on a quantitative basis, by demonstrating that the brightness variation in the schlieren image was related to the refractive index gradient in the disturbance. Since then schlieren methods have found use wherever refractive index gradients are present. The widespread application of the technique has resulted in the development of many variations, from simple single concave mirror systems to colour and laser illuminated holographic systems. A major advantage of the schlieren method and optical techniques in general is that they are non-invasive i.e. there is no physical interaction with the inhomogeneity under study. Conventional hydrophone measurements often suffer from limited time response and produce interference effects due to the reflection of the incident acoustic wave. A short bibliography and review of the various schlieren techniques and system configurations can be found in a publication by Davis.

One of the best known schlieren techniques is that due to Toepler, in which a knife edge is used to spatially filter light passing through the working section of the system. With no disturbance in the working section, the system is set such that the light is partially cut off by the knife edge, resulting in the entire image plane of the system being dark. This is not so when an inhomogeneity is present, since the light is now deflected by the density gradient of the disturbance and bypasses the knife edge. The image of the inhomogeneity thus appears illuminated against the dark background of the image plane. One of the simplest methods of observing optical disturbances is the shadowgraph method. Originally proposed by Dvorak, the method requires no spatial filtering, instead undeviated light passes through the working section and illuminates the image plane. Any light deflected by an inhomogeneity thus forms a shadow or dark region in the image. If the deflected light remains in the system a bright region is also
formed adjacent to the region of shadow. Theory shows that the change in illumination in the image plane of the schlieren and shadow systems, is proportional to the density gradient and the rate of change of the density gradient respectively. The shadowgraph method is therefore unsuitable for the study of slow and smooth changes in density and is thus principally used for the study of large density gradients. Observation of these slow changes in density can be made using the more sensitive schlieren method.

In many cases quantitative as well as qualitative information about an inhomogeneity is required. One of many ways of achieving this in a conventional schlieren system is to change the level of spatial filtering. Often this takes the form of a series of photographs taken with different knife edge settings. By simply comparing the required knife edge displacement to block a particular component of the image, the angular deflection and thus density gradient can be estimated. One of the drawbacks of the technique is that a number of photographs are required, this method is therefore unsuitable for the study of unsteady disturbances. Defocused grid methods on the other hand allow the simultaneous measurement of the entire deflection field of an inhomogeneity. This, coupled with the simplicity with which the image can be interpreted, makes it particularly suitable for the analysis of the complex acoustic shock fields generated by the EMAT. In the following chapter the theory behind the deflection of a ray of light by a density gradient and its observation using schlieren techniques will be discussed. The chapter will also discuss the effects of diffraction on the schlieren image and the synchronisation and timing details of the key components in the system as well as the set up procedure for the system as a whole.

4.2 The relationship between refractive index, density and pressure

The refractive index $n$ of a substance is a purely optical characteristic and depends to a large extent on the density and temperature of a substance and the wavelength of the incident light. The effect of density is highlighted here for water and water vapour, where $n$ ($\lambda = 694$ nm) can be seen to vary considerably depending on whether it is in its vapour ($n = 1.00025$) or liquid state ($n = 1.333$). It can be anticipated that the variation in density is the main factor causing the change in $n$. The exact mechanism by which a change in density leads to a change in $n$ is beyond the scope of the present discussion, suffice to say that it gives rise to a change in the polarisation of the medium and or the polarisability of the individual molecules.
In an earlier discussion the propagation of an acoustic wave of pressure $P$ was found to lead to an increase in the local density. From the above, it can be seen that this also leads to a corresponding increase in the local refractive index of the medium. The relationship between refractive index and the density $\rho$, can be expressed independently of temperature by any one of a number of expressions, such as the Gladstone-Dale\textsuperscript{6}, the Lorentz-Lorenz\textsuperscript{7,8}, the Newton\textsuperscript{9} and the Eykman\textsuperscript{10} formulae. Not all of the aforementioned expressions fit direct observations of substances; however, the Gladstone-Dale and Lorentz-Lorenz expressions are usually both satisfactory when considering the compression of liquids\textsuperscript{6,11-13}. According to the independent studies of Lorenz and Lorentz, the refractive index is related to the number of molecules per unit volume, expressed in terms of density the relation can be written thus.

$$\frac{n^2 - 1}{n^2 + 2} \propto \rho \quad (4.01)$$

For gaseous like media where $n$ is very nearly equal to unity, the above relation can be simplified to give the Gladstone-Dale relation,

$$\frac{n - 1}{\rho} = k \quad (4.02)$$

where $k$ is a constant for a particular wavelength of light. Recent work by Vedam\textsuperscript{14} on the variation of the refractive index of water with pressure up to 1.4 GPa, has shown that $n$ increases non-linearly with pressure, particularly at high pressures when it begins to tail off. His results for lower pressures are in excellent agreement with similar measurements on water made by Waxler & Weir\textsuperscript{15} and Rosen\textsuperscript{16}. In most cases the Gladstone-Dale model has been found to produce acceptable results to around 500 MPa. At larger compression's Zeldovich’s\textsuperscript{17} temperature dependent Gladstone-Dale model is more suitable.

If the values of $\rho_o$ and $n_o$ are known at a certain standard temperature and pressure, it is convenient to write equation 4.02 in the form,

$$\frac{n - 1}{\rho} = \frac{n_o - 1}{\rho_o} \quad (4.03)$$

from which it can be shown that,
\[ \frac{dp}{dn} = \frac{\rho_o}{n_o - 1} \]  

(4.04)

In addition, Willard\textsuperscript{18} states that

\[ \frac{dp}{\rho_o} = \beta_a dP \]  

(4.05)

where \( \beta_a \) is the adiabatic compressibility and \( P \) is the hydrostatic pressure. Therefore, given that,

\[ \beta_a = \frac{1}{v^2 \rho_o} \]  

(4.06)

equation 4.05 can be written thus,

\[ \frac{dp}{\rho_o} = \frac{1}{v^2 \rho_o} dP \]  

(4.07)

where \( v \) is the speed of sound in the medium. Combining equations 4.04 and 4.07, we obtain an expression that relates the change in refractive index to the pressure.

\[ \frac{dn}{dP} = \frac{(n_o - 1)}{v^2 \rho_o} \]  

(4.08)

Evaluating (8) for \( n=1.33 \) (\( \lambda=694 \text{ nm} \)) gives \( \frac{dn}{dP} = 1.50 \times 10^{-4} \text{ (MPa)}^{-1} \). Based on this relationship, a graph of the refractive index of water against a range of pressures from 0-100 MPa has been plotted in figure 1.

Measurements on the temperature dependence of the refractive index for water by Waxler\textsuperscript{15}, has shown that \( n \) decreases with increasing temperature. At a wavelength of 694 nm the change in \( n \) with respect to the temperature \( T \) \( (dn/dT)_{\rho=\text{const}} \) is estimated to be \( -6 \times 10^{-5} \text{/°C} \). This effect is found to increase in many liquids for light wavelengths in the violet end of the spectrum. However, since in this case the temperature rise associated with the shock compression of the water is around 0.2 °C, the change in the refractive index can be assumed due totally to the change in density.
Chapter 4: Schlieren methods

4.3 The Toepler schlieren method

The observation of an optical inhomogeneity in water, where the only physical sign of its presence is a change in density, presents obvious problems for conventional photography. Fortunately, as we have already seen such changes in density are accompanied by a corresponding change in the refractive index. Techniques such as schlieren photography that are sensitive to these changes can therefore provide a suitable means of observing such phenomena. If a density gradient exists such that there is a refractive index gradient in a direction normal to the incident light rays, the rays will be deflected. The direction of the deflection will be towards the refractive index gradient i.e. towards the region of highest density, as light travels slower in regions of higher refractive index. The theory behind the bending of light and the relationship between the refractive index gradient and deflection angle, is discussed with the help of Huygens principle for a planar inhomogeneity in appendix III.

The theory shows that the curvature of a light ray is proportional to the refractive index gradient in a direction normal to the ray. If the z-axis is taken to be the direction of the undisturbed light ray, then the deflections $\varepsilon_x$ and $\varepsilon_y$ of the ray in the x-z and y-z planes are given by\textsuperscript{19}.
Chapter 4: Schlieren methods

\[ \varepsilon_x = \int_{z_1}^{z_2} \frac{1}{n} \frac{dn}{dx} \, dz \]

\[ \varepsilon_y = \int_{z_1}^{z_2} \frac{1}{n} \frac{dn}{dy} \, dz \]

(4.09)

where \( z_1 \) and \( z_2 \) are the entrance and exit points of the ray in the inhomogeneity. Figure 2 shows a typical two lens parallel beam schlieren arrangement suitable for the qualitative and quantitative analysis of acoustic shock phenomena. In this simplified diagram light from the ruby laser source is first collected and spatially filtered, before being expanded and collimated. The parallel beam of light then passes through the working section of the schlieren system, prior to being focused by the schlieren lens to form an image of the source in the focal plane. A further lens positioned beyond the focal plane forms a conjugate image of the working section in the image plane of the schlieren system.

The knife edge is positioned in the focal plane such that it cuts off part of the light forming the source image. Therefore, under normal conditions with no disturbance present in the working section, the level of illumination in the image plane of the system will be low but uniform. However, if an optical disturbance is introduced such that a component of the refractive index gradient is normal to both the optical axis and the knife edge, the light rays will be deflected. The angular deflection \( \varepsilon \) of the light will cause the corresponding image of the source in the focal plane to be moved. The displacement of the source image being approximately equal to \( f_2 \varepsilon \) for small angles, where \( f_2 \) is the focal length of the schlieren lens. Thus light passing through a particular point in the working section, results in a change in the illumination of the
image of that point in the image plane. In the case of the schlieren method, this change in illumination is proportional to the refractive index gradient. The respective darkening or brightening of a particular part of the image, will depend on the direction of the deflection and whether the light falls on or away from the opaque side of the knife edge. Light deflected parallel to the knife edge produces no change in illumination, the edge must therefore be set perpendicular to the gradients to be observed. To observe these rays, the knife edge would have to be turned through 90 degrees. The sensitivity to the deflection direction and thus particular density gradients, allows knife edge systems to be arranged so that only those features of greatest interest are observed with high sensitivity. For general observation, it is often more convenient to have a system which is sensitive to density gradients in both directions and both planes. In such systems the knife edge can be replaced by a spatial filter in the form of a circular hole or an opaque circular stop. However, an important point to bear in mind when using such filters, is that the spatial period of the bright or dark regions in the schlieren image is now half that of the pressure wave producing the distortion. In cases where the overall illumination is severely limited or the inhomogeneity is self luminous the circular hole filter is often the best choice, since any displacement of the source image results in decreased illumination.

The use of schlieren methods to obtain quantitative data such as the value and position of the gradients that give rise to the observed image, often requires a lot of prior knowledge of the inhomogeneity. This is particularly so with 3-dimensional inhomogeneities, where the density varies across the working section along the path of the ray. The main reason for this is that it is the total angular deflection of the ray and not the behaviour of the ray in a particular part of the inhomogeneity that is measured. It is therefore quite possible that a light ray may suffer the same total deflection and yet have been subjected to quite different conditions. In fields that are 2-dimensional, in which the density is assumed constant across the inhomogeneity or that possess axial symmetry, the relationship between the total deflection and density gradient can often be determined. The important problem is therefore being able to measure the total angular deflection of a light ray accurately. Methods of doing this will be discussed shortly.

4.31 The focused shadow method

The versatility of the schlieren method is due in part to the large variety of techniques available. One of the simplest and least expensive is the direct shadow method. In its
most basic form a light source, a collimating spherical mirror or lens and a photographic plate or screen is all that is required. Although, it is more usual to modify a conventional schlieren set up such as that shown in figure 2, by removing the stop in the focal plane of the system. As with the schlieren method described earlier, density gradients brought about by the presence of a disturbance in the working section, gives rise to the deflection of light rays in the system. However, unlike the schlieren method a disturbance with a uniform density gradient produces no change in illumination in the image plane, since all the rays are deflected by the same angle. The shadow method is therefore only sensitive to those inhomogeneities with non uniform density gradients, in which rays are deflected by different amounts. The change in illumination in the image plane is thus proportional to the rate of change of the density gradient, i.e. the second derivative of the refractive index gradient. Hence, shadowgraphy is particularly suited to the study of density discontinuities such as those that occur with shock phenomena, but ill-suited to studying slow and smooth changes in the density of a medium.

Since the inhomogeneities typically observed with this type of system produce large deflections, it is often necessary with the direct method to position the screen near to the working section, to prevent the loss of that information to the system. This can sometimes be inconvenient, so in these cases the focused shadow method can be used. The system, generally based on the arrangement of figure 2, minus the stop, uses the schlieren lens (as with the schlieren method) to produce a conjugate image of the working section in the image plane. However, in addition to the system having a finite depth of focus, the aperture of the lens ultimately acts as a stop, passing all but the most strongly deflected or diffracted light. As a result the image of the disturbance is often different to that obtained using the direct method. The principle behind the focused shadow method is illustrated in figure 3.

![Figure 3 The principle of the shadowgraphic method.](image)
In the diagram the three points a,b and c, represent the different parts of a non uniform inhomogeneity in the working section of the shadow system. Rays from 'a' collimated beam of light passing through these regions, undergo a deflection depending on the strength of the gradient at that point. In this case the region surrounding point 'b' is assumed to be the position of the shockfront and therefore produces the largest deflection. Region 'c' is in the low pressure tail of the shock wave and as a result is shown to produce little or no displacement. Assuming the rays remain in the system and are captured by the lens, the image of the inhomogeneity can be seen in the image plane. The image of the shock wave will consist of two adjacent bands, one bright, one dark. The dark or shadowed region will be visible between points a'c' and is associated with the deflection of light away from this region by the shockfront. the bright band produced by the deflected rays will be imaged at point b' outside region a'c'. Quantitative information about the inhomogeneity can be obtained from the shadow images by measuring the width of the shadow region\textsuperscript{23}, however, this is seldom undertaken due to the low sensitivity of the method.

4.32 The scale method

The scale method\textsuperscript{24} belongs to a group of quantitative schlieren techniques known as defocused grid methods. The main characteristic of this group is the use of a grid of opaque spots, parallel lines, concentric circles, etc. to determine the angular deflection of light in the working section of a schlieren system. The individual methods differ only in the location of the grid, which in the case of the scale method is positioned before the inhomogeneity in the object plane of the system. This makes the scale method particularly suited to the study of inhomogeneities that deflect light through large angles. As with schlieren knife edge systems, by the choice of grid type, grid methods can be made to image only particular density gradients in the inhomogeneity. For instance, a mutually perpendicular system of lines can simultaneously recover the entire deflection field of an inhomogeneity. Whilst, a parallel set of lines is sensitive only to deflections perpendicular to that particular system, c.f. knife edge. One of the major advantages of grid methods is that object illumination need not be uniform since, it is grid displacement not illumination that is measured. The level of illumination need only be sufficient to resolve individual line displacements.

The presence of an optical disturbance such as a shock wave propagating through the working section, causes a displacement of the grid lines in the image plane of the system. The displacement of a point on the grid is proportional to the refractive index
gradient $dn/dx$, associated with the pressure profile of the shock wave at that point. The accuracy with which the displacement can be measured is not effected by the grid line spacing, although too few or too many grid lines can prevent full analysis of the inhomogeneity. Figure 4 shows a simple representation of the path of a ray from a point on the grid, traced through the schlieren system. A collimated laser beam in the plane of the page enters the system from the left along the z-axis and illuminates a shock wave propagating along the x-axis, through the working section of the schlieren system. The optical axis is denoted by the line AA and the working section as the region between the planes $P_2$ and $P_3$. The image of the scale at $P_1$ is focused by the schlieren lens $P_4$ onto the image plane $P_5$.

![Figure 4 A schematic diagram of the scale method.](image)

Referring to figure 4, let us suppose a point in the object plane $P_1$ is ray traced through the system. In the absence of the shock wave in the working section, the image of this point will occur at its conjugate position I in the image plane $P_5$. This is not the case when a shock wave is present. In its passage through the inhomogeneity the ray will be deflected, resulting in an angular change of $\epsilon$ from its undeviated position with no shock present. Tracing this ray back through the inhomogeneity gives the apparent position $s'$ of the point $s$ on the object plane and therefore its position in the image plane $I'$. Associated with the displacement is a degree of defocussing due to the focal length of the particular part of the shock wave through which the ray has just passed. A region of compression places the point $s'$ behind the object plane as indicated in figure 4, whilst a region of rarefaction locates the point in front of the object plane. Both examples result in the image of the point on the grid or scale being blurred. For small angular deviations the displacement $\delta$ of the grid lines from I to $I'$ is given by,

$$\delta = Mb\epsilon$$

(4.10)
where $\varepsilon$ is the deviation angle, $b$ the distance between the scale grid $P_i$ and the centre of the working section $P_0$ and $M$ the system magnification. The relationship between the deviation angle $\varepsilon$ and the refractive index gradient $dn/dx$ from appendix III is,

$$\varepsilon = \frac{dn}{dx} \Delta z$$

(4.11)

where $\Delta z$ is the path length $(z_2 - z_1)$ through the shock wave. In the distance $\Delta z$, $dn/dx$ is assumed to have a constant value. Combining equations 4.10 & 4.11 relates the grid line displacement $\delta$ in the image plane to the refractive index gradient produced by the shock wave in the working section of the schlieren system. The grid line displacement $\delta$ is therefore given by,

$$\delta = Mb \frac{\Delta n}{\Delta x}$$

(4.12)

where $\Delta n/\Delta x \equiv dn/dx$, as $dn$ and $dx$ are small.

### 4.4 Quantitative analysis using knife edge systems

The main drawback of using the scale method for the qualitative and quantitative analysis of acoustic shock phenomena, is that only a discrete set of ray deflections can be measured. As a result the determination of the true shape and character of the inhomogeneity can become more complicated. In the Toepler schlieren method described in section 4.3, it was seen that an inhomogeneity in the object plane can lead to a change in the image plane illumination. Assuming the maximum ray displacement lies within the working range of the schlieren system, the level of illumination will be related to the displacement. By recording the image illumination using photographic film and then measuring the density of the negative, it is possible in principle to determine the angle of deviation. Methods that rely on this technique are known as photometric methods\textsuperscript{25}. Included in this group is a technique in which standard schlieres such as a glass wedge\textsuperscript{3} or bubbles\textsuperscript{26}, with known angular deflections are positioned in the field to act as a reference point for density measurements. The difficulties involved in accurately determining the density and relating this to the angular deflection of a ray can be considerable. This is particularly so if diffraction effects, which will be discussed shortly, become significant or the working section is not uniformly illuminated.
An alternative approach is to take a number of photographs of the inhomogeneity with different knife edge settings. From the images, the displacement of a number of rays associated with particular characteristics of the schlieren image can be calculated. The advantage of this approach is that ray displacement measurements are less sensitive to the illumination changes introduced by diffraction and non uniformities in the source illumination. The use of a knife edge spatial filter to obtain quantitative information, results in images that contain only part of the object information at a particular sensitivity. This is because the knife edge will selectively pass only positive or negative values of the refractive index gradient $dn/dx$; that is, light deflected to the right or left of the spatial filter. By noting the required knife edge displacement to extinguish a particular part of the image, a figure for the angular displacement of the rays can be calculated. From this it is then possible to estimate the density gradient and thus the pressure of a shock wave propagating through the working section of the schlieren system.

From equation 6 appendix III, the angular deflection of a ray in the x-z plane is given by,

$$\varepsilon_x = \int_{z_i}^{z_f} \frac{1}{n} \frac{dn}{dx} dz$$

(4.13)

The total deflection due to an inhomogeneity can therefore be approximated by the sum,

$$\varepsilon_x = \sum \varepsilon_{x_i} = \sum \frac{1}{n_i} \frac{\Delta n}{\Delta x} \Delta z_i$$

(4.14)

where $n_i$ is the refractive index at a certain point in the inhomogeneity, separated from the next point by a finite distance $\Delta z_i$. Rays deflected by the refractive index gradient, are spatially filtered in the Fourier transform plane of the schlieren lens (focal length $f_2$) by a knife edge of width $r$ or an opaque disc of radius $r$. An opaque disc will extinguish all rays with a deflection $\varepsilon$ in the interval $-\varepsilon < 0 < \varepsilon$, thus for small angles the deflection is given by,

$$\varepsilon = \tan^{-1} \frac{r}{f_2} \approx \frac{r}{f_2}$$

(4.15)

By assuming the refractive index gradient is constant over a distance $\Delta z$ and transforming equation 4.14, it is possible to determine the refractive index gradient of
a component within the schlieren image from its deflection angle \( \varepsilon \). From equations 4.14 & 4.15 we can write.

\[
\frac{\Delta n}{\Delta x} = \frac{n_f}{f_2 \Delta z_i}
\]  

(4.16)

Thus by considering equation 4.08 it is possible to obtain an approximation for the pressure of a 2-dimensional inhomogeneity from the associated ray deflection.

\[
\frac{\Delta P}{\Delta x} = \frac{n_f r^2 \rho_0}{f_2 (n_0 - 1) \Delta z_i}
\]  

(4.17)

Where \( \Delta x \) is the average rise distance i.e. the rise time of the wavefront multiplied by the speed of sound in the medium.

4.41 The photographic process

Photographic film remains one of the most important recording mediums in use today. Alternatives, such as charge coupled devices (CCD's) are developing quickly and although image processing time is substantially reduced, they are still unable to compete on cost and simplicity. In most cases all that is required to record a schlieren image with conventional photography, is a basic camera and a source of illumination. However, for transient disturbances where high temporal resolution is required, short duration arc or pulsed laser sources may be necessary. The importance of obtaining a good recording of an image is seen where, the data is to be processed at a later date. Quality is therefore essential. To achieve this an understanding of the characteristics and physical process involved in photography, particularly if conventional film is to be used, is very helpful.

An unexposed black and white film generally consists of a multitude of light sensitive silver halide grains suspended in a gelatine layer\( ^{28} \). This emulsion is in turn attached, in the case of a transparency, to a base of cellulose acetate. When the photosensitive emulsion is exposed to light, the silver halide grains breakdown to produce patches of metallic silver (known as development centres) and bromine. During the development process, these centres precipitate the change of the entire silver halide grain into metallic silver. Those grains unexposed remain unchanged. Fixing, then chemically removes the remaining silver halide whilst leaving the opaque metallic silver, preventing any further breakdown. The image on the developed transparency,
therefore, consists of opaque regions of silver whose opacity is dependent on the density of silver grains in that region. If \( I_0 \) is the intensity of the incident light on the film and \( I_t \) the transmitted intensity, the opacity is simply \( I/I_t \) and the density \( D \) the log of opacity.

\[
D = \log_{10} \left( \frac{I_0}{I_t} \right) \quad (4.18)
\]

From theory, the exposure \( E \) is defined as the energy per unit area incident on the photosensitive surface and is given by,

\[
E = \int_{t_1}^{t_2} I(t) \, dt \quad (4.19)
\]

where \( I \) is the incident illumination and \( t \) the duration of exposure. The relationship between exposure and the density developed under given conditions for an emulsion, can be expressed by a curve or series of curves, commonly referred to as Hurter-Driffield (H&D curves). A typical H&D curve, in which photographic density is plotted against log \( E \) is shown in figure 5.

The curve shows that the characteristic of an emulsion consists of three parts, the first is an underexposed region, referred to as gross fog, where the density is independent of exposure. This is followed by the toe and the linear part of the characteristic, where the density is proportional to log exposure and then the shoulder or the overexposed region, where the curve again flattens out. The tangent of the angle \( \alpha \), which the linear part of the characteristic makes with the log \( E \) axis is known as the film gamma (\( \gamma \)). This is a measure of the contrast of the film, i.e. the rate at which density grows with increasing exposure. The relationship between the film \( \gamma \), which rises to a limiting value known as gamma infinity, and development time is shown inset figure 5. As well as development time the particular value of \( \gamma \) is also dependent on two other factors, namely the type of emulsion and the type of developer used. High contrast films have gamma's of between 2 or 3, whilst low contrast films have gamma's of less than 1. In most cases it is the linear part of the films characteristic that is used, this is in order to maintain a proportional relationship between the object brightness and the negatives density. In such cases the density may be written thus,

\[
D = \gamma \log_{10} E - D_o \quad (4.20)
\]
where \( D_0 \) is the value of \( D \), obtained if the linear part of the characteristic were continued to meet the \( D \) axis. Films that are underexposed often require larger periods of development to obtain shadow detail. As a consequence fogging and grain can increase, resulting in the negative appearing grainy and hard.

![Figure 5 The Hurter-Driffield curve.](image)

The production of a negative is only the first stage of the photographic process. In the second stage, light transmitted through the negative transparency is used to expose light sensitive paper, producing the final positive image. The relationship between the contrast of a print and the original schlieren image is given by,

\[
C_p = \gamma_p C_n = \gamma_p \gamma_i C_i
\]  

(4.21)

where \( C \) is the contrast and the subscripts \( i, n, \) and \( p \) represent the image, negative and the print respectively.

### 4.42 Sensitivity

The sensitivity of a measuring instrument is defined as the ratio of the incidental change in the instrument reading to the incidental change in the value being measured. Thus for a schlieren system, where the fixed parameter is the optical density \( D \) of the photo emulsion and the measured parameter the ray deflection angle \( \varepsilon \), the sensitivity \( S \) can be written thus.

\[
S = \frac{dD}{d\varepsilon}
\]  

(4.22)
For the linear part of the film emulsion characteristic (H&D curve) the optical density is related to the illumination $I$ by equation 4.20, hence the sensitivity can be rewritten thus:

$$S = \gamma \frac{dI}{I \, d\varepsilon} \quad (4.23)$$

From theory the illumination $I_o$ in the image plane, in the absence of the knife edge and neglecting losses is given by,

$$I_o = \frac{Bwh}{M^2 f_1^2} \quad (4.24)$$

where $B$ is the luminance, $w$ is the width and $h$ the height of the light source, $f_1$ the focal length of the first lens (see figure 2) and $M$ is the system magnification. If all but a width $a'$ of the source is now cut off by the knife edge in the focal plane, in which the source dimensions are multiplied by a factor $(f_2/f_1)$, the illumination $I$ falls to,

$$I = \frac{Bha}{M^2 f_1 f_2} \quad (4.25)$$

where $f_2$ is the focal length of the schlieren lens. Thus the change in illumination associated with the displacement of the source image, given with sufficient accuracy by $f_2\varepsilon$ for small angles, by an inhomogeneity in the working section is given by.

$$dI = \frac{Bh\varepsilon}{M^2 f_1} \quad (4.26)$$

The contrast $C$ of the image with respect to the back ground is therefore,

$$C = \frac{dI}{I} = \frac{f_2\varepsilon}{a} \quad (4.27)$$

and the contrast sensitivity $S$ from equation 4.23 is thus.

$$S = \frac{\gamma f_2}{a} \quad (4.28)$$

Equation 4.28 suggests that the contrast sensitivity can be increased without limit by increasing either the focal length of the schlieren lens or reducing the width $a'$ of the
source. In practice however, sensitivity is ultimately limited by the effects of diffraction and aberration, which become significant at high sensitivities and the reduction in sensitivity of the photographic emulsion at low levels of illumination (c.f. H&D curve). The maximum angular deflection $d\varepsilon$ and thus working range over which this sensitivity can be obtained is given by the following.

$$d\varepsilon_{\text{max}} = \frac{w}{f}$$  \hspace{1cm} (4.29)

Inhomogeneities producing ray deflections greater than $d\varepsilon_{\text{max}}$ will result in the source image being displaced completely on, or completely off the opaque side of the knife edge. Consequently, the image plane will either be totally dark or have an illumination equal to $I_0$ (Eq.4.24) and as such be insensitive to any further displacement. Under these conditions equation 4.28 no longer applies. The practical implication of exceeding the working range of the system, is that there may be little contrast between the various components of the image. In such circumstances it may be more appropriate if quantitative information is required, to measure ray displacement by either changing the level of spatial filtering or by the use of the scale method. In both cases the sensitivity $S$ is redefined thus,

$$S = \frac{d\delta}{d\varepsilon}$$  \hspace{1cm} (4.30)

where $\delta$ is the ray displacement. For the spatial filter method, this is measured in the focal plane of the schlieren system and given by $f_2\varepsilon$. Hence from equation 4.30 the sensitivity is,

$$S = \frac{f_2\varepsilon}{\varepsilon} = f_2$$  \hspace{1cm} (4.31)

here $f_2$ is the focal length of the schlieren lens. For the scale method, the displacement of grid lines in the image plane is given by equation 4.10. The sensitivity of the method is therefore given by,

$$S = \frac{Mb\varepsilon}{\varepsilon} = Mb$$  \hspace{1cm} (4.32)

where $M$ is the system magnification and $b$ is the distance between the grid in the object plane and the centre of the working section.
4.5 Optical considerations

In the literature schlieren schemes can be found in abundance. They typically consist of arrangements of either mirrors or lenses configured in such a way as to make use of their relative properties. Mirror systems employing one or two concave mirrors in parallel or non parallel, single or double pass beam systems, form some of the simplest and most cost effective systems available. For quantitative analysis single pass parallel beam systems, where the working section is illuminated by a parallel beam of light are the most desirable. Although, in certain circumstances where increased sensitivity is required, double pass systems in which a beam crosses the working section twice may be more appropriate. Schlieren systems using non parallel light i.e. a single lens or mirror system, are usually restricted to qualitative analysis due to uncertainties over the dimensions of the beam working area.

The cost effectiveness of mirror systems in comparison to those based on lenses, stems from the less stringent optical requirements for mirrors. The only other requirement of a mirror that has been figured to minimise aberrations, is that the surface be free from material defects. This is in sharp contrast to lenses, where along with optical errors introduced by chromatic and spherical aberrations, coma and astigmatism, require the lens material to be of high quality and free from internal flaws. These are commonly in the form of seed (gas bubbles), stone (solid inclusions) or ream (streaks or striations of glass whose refractive index differs appreciably from the surrounding glass). The cost of large diameter lens based systems can therefore be prohibitive unless, as in this case military surplus aerial telephoto lenses can be obtained. Other system components, such as the windows that form the working section of the acoustic tank must also adhere to the stringent requirements. In addition to those internal defects mentioned above, the windows must also be flat and positioned parallel to each other. Although a lack of parallelism, which leads to the uniform deflection of the beam, can normally be corrected by adjusting the knife edge. Another particularly important component of the schlieren system is the light source. Generally the choice will depend on the application and the schlieren method used. However, the most important properties are usually its brightness, dimensions and duration. The latter is of particular importance if high speed shock wave phenomena are to be observed. Pulsed laser sources can usually satisfy most of these requirements and being monochromatic eliminate chromatic aberrations in the optical components.
4.51 Diffraction

The geometric theory used so far to describe the deflection of a ray by a density gradient and the resulting change in the image plane illumination, neglects one important phenomena, the wave theory of light. Under certain circumstances, as a result of diffraction the image plane illumination associated with an inhomogeneity can be substantially different to that expected under geometrical optics. Whilst, as with geometric theory, the image plane illumination in the absence of a spatial filter may still be approximately uniform, this will not be the case when the filter is present. In most cases a sudden increase in the illumination at the boundary of the system and around the edges of any object or wavefront in the working section will be visible. As the spatial filter (knife edge) is advanced and an increasing proportion of the source image is cut off, the magnitude and extent of these diffraction effects increase. Around the boundaries, the illumination no longer falls sharply to zero, but consists of a number of bright interference fringes running adjacent to them. The influence of diffraction in an image will depend on the schlieren method, its operating sensitivity, the type of spatial filter and the shape and characteristics of the object or wavefront. In a conventional schlieren system (figure 2) employing a knife edge, the diffraction fringes will only be visible on those edges of an object or wavefront that are parallel or have a component parallel to the knife edge. Fringes on edges perpendicular to the knife edge will not be visible. An opaque spot spatial filter on the other hand, will yield the complete diffraction field. Because diffraction effects cannot be eliminated, they ultimately determine the upper operational limits for accuracy and sensitivity of schlieren methods. However, when used at low sensitivities for the observation of inhomogeneities producing large ray displacements, diffraction does not unduly effect qualitative or quantitative analysis. The comprehensive analytical analysis of the effects of diffraction on image plane illumination in schlieren systems can be found in texts by Speak and Vasilev.

4.6 Experimental apparatus

An essential requirement of any experimental work is the need for consistency, particularly when generating the results. The method of generation and the nature of the shock wave, meant that achieving reproducible results was always expected to be difficult. One of the main reasons for this is the variation in the operational characteristics and the reliability of the apparatus. Operational reliability was therefore identified from a very early stage as being of prime importance, particularly since most
of the key EMAT components were designed from first principles, within a short time period and on a limited budget. The EMAT and its associated drive electronics, together with the hydrophone described in chapter 3 and the schlieren imaging techniques described earlier in this chapter, follow a simple and rugged design philosophy to ensure maximum reliability. The integration of the EMAT and the schlieren apparatus followed this same principle and provided the means of photographing the shock waves for the first time. The better than expected results obtained from the system allowed many aspects of shock wave propagation in liquids to be studied.

4.6.1 The system layout

The experimental arrangement used for the optical characterisation of acoustic shock waves generated by the EMAT is shown in the simplified schematic of figure 6. The main components of the system are the EMAT, its test tank and the basic schlieren system, which intersects the acoustic field in a region of the tank known as the working section. The tank, apart from one machined aluminium end section that supports the EMAT, is constructed from high quality float glass. Its dimensions are 600x300x300 mm and is filled with distilled water to limit the impurity content of the water and help reduce corrosion. Within the working section, planar and focused shock waves generated by the EMAT deflect a schlieren probe beam to produce the schlieren images described earlier. The generation of these shock waves follows the capacitor discharge of approximately 88 J (depending on charge voltage, usually 20 kV) through the EMAT coil. The resulting mechanical displacement of the EMAT diaphragm radiates a planar acoustic transient with a maximum pressure of 5 MPa into the surrounding medium. Variation between the shots was less than 5% of the plane wave pressure. The addition of the focusing acoustic lens increases the maximum pressure to around 20 MPa or 200 bar at the focal point.

The schlieren configuration is a conventional two lens parallel beam system with additional optics for beam expansion and image de-magnification. Illumination for the system was provided by a JK lasers system 2000, 20 ns duration Q-switched ruby laser, operating at 694 nm. To illuminate the working section of the schlieren system, it was first necessary to expand and then collimate the laser beam to 100 mm in diameter. The expansion optics comprised of a 16 mm focal length bi-convex lens with a second lens f=250 mm f/2.5, placed at a distance equal to the sum of the focal lengths apart. At the common focal point of the combination, a 100 μm pinhole spatial
filter was used to filter out random spatial noise from the beam. The collimated beam then passes through the tank and illuminates a 100 mm diameter cylindrical region of the working section, which is oriented 90 degrees to the EMAT axis and centred on the focal zone of the acoustic lens. Since the acoustic beam is also confined within a cylindrical region of approximately 100 mm in diameter, the actual active area in the working section is around 100x100x100 mm.

![Diagram of experimental arrangement](image)

Figure 6. The experimental arrangement for the study of acoustic shock waves using schlieren photography.

After traversing the working section, a schlieren lens (a war surplus Avia camera lens $f=508$ mm $f/5$) placed at $2f$ from the working section brings the beam to a focus at $f$, where it is spatially filtered with the schlieren stop. For general observation this was an opaque disk, but was replaced by a knife edge for the quantitative studies, allowing adjustment to the level of spatial filtering in the system. A fourth lens ($f=140$ mm) was positioned in the system to form a de-magnified image (~5x) of the working section in the conjugate image plane of the system. The images were recorded using a conventional 35 mm camera and Ilford FP4 black and white film, with a shutter speed of 1/125 of a second. As no other light was able to enter the system, the exposure time was effectively determined by the laser pulse duration. This gave the system a spatial resolution of approximately 30 μm, for a shock wave speed in water of 1480 ms$^{-1}$. By controlling the laser timing to illuminate the shock waves at different stages in their development during their passage through the working section, schlieren and shadow images of the shock waves were obtained.
The maximum sensitivity that can be obtained from the system depends on the schlieren method used. For the photometric method (Equ. 4.28) where the width of the open part of the slit 'a' is around 1x10^-6 m and the emulsion contrast factor γ (FP4) is 0.55, the contrast sensitivity S is 279x10^3. Assuming that the maximum film contrast dI/I is approximately 10%, then at maximum sensitivity the system is capable of resolving deflection angles of around 1.97x10^-7 radians. For the knife edge (Equ. 4.31) and scale method (Equ. 4.32), the sensitivity is calculated to be substantially lower at 508x10^-3 and 198x10^-3 respectively. Therefore, assuming that the minimum measurable ray displacement in both cases was around 0.1x10^-3 m, we can expect to resolve deflection angles of 1.97x10^-4 radians for the knife edge method and 5.05x10^-4 radians for the scale method.

4.611 Laser illumination

Previous discussions have highlighted the importance of the right source of illumination for a schlieren system. Where fast moving shock phenomena are to be observed, one of the most important characteristics is pulse duration. Lasers are often the most convenient source of short duration illumination, although the pulse length of such sources can vary widely. The original source used with the prototype of the system described earlier (figure 6) was a dye laser. The development laser system utilised an optically pumped dye cell (Rhodamine 6G) and operated at a wavelength of 514 nm. Whilst in most respects the laser was satisfactory, the pulse length which was around 2 μs, was unacceptably long and resulted in the images of the shock wave being blurred. In contrast to the ruby which later replaced the dye laser, its pulse length corresponded to a spatial resolution of approximately 3 mm, three orders of magnitude larger. The superior spatial resolution of the ruby was complimented by a second feature which allowed the laser to operate in a double pulse mode. In this mode, two pulses separated by a delay of between 1-10 μs could be selected by the operator. This facility was particularly useful in the study of cavitation shock transients. Figure 7 shows the shape of the laser pulse generated while operating in the standard, single pulse mode.

The trace shows that the pulse is characterised by a steep rising edge, followed by a much slower falling edge, which is almost certainly associated with the characteristics of the photodiode rather than a property of the laser radiation. Another important characteristic of the laser beam was its spatial uniformity, which in this case was far from satisfactory due to a defect in the laser's synthetic ruby (Al₂O₃) laser rod. Prior to
its repair, to obtain the desired beam quality the fault necessitated operating the laser at near its maximum operating voltage and intensity. This brought about its own problems, one of which was film overexposure due to the intensity, although this was rectified by attenuating the beam using several neutral density filters.

![Figure 7](image)

At 694 nm, the laser illumination was at one end of the film's spectral sensitivity. Being primarily used for conventional photography, the film's sensitivity to this wavelength was lower than that for white light. Exposure to white light, particularly that entering the system from the room was therefore avoided by dimming the room lights during the experiments. However, one other source, which entered the system as a component of the laser beam and produced by the two laser flash tubes, required the use of a single 680 nm high pass filter.

4.6.12 Apparatus set-up procedure

The first stage of the set-up procedure required the ruby laser source and the optical rails to be aligned. This was initially done with a continuous HeNe laser, since the short pulse duration and low repetition frequency of the ruby laser made it unsuitable for alignment purposes. The HeNe beam which could be switched in and out of the system using a removable mirror, was aligned as near as possible to the beam axis of the ruby. As a result it was only necessary to use the ruby to fine tune the adjustment of the optical components in the system. Prior to entering the tank, the beam was collimated using the beam expansion and collimation optics described earlier. To ensure the beam was correctly collimated and remained aligned with the axis of the system, the beam position and size was checked at several points along the optical rail. Any skew introduced by the optics or lack of collimation in the beam, was removed by making the necessary adjustments. Beam quality was also monitored, particularly the
uniformity of the illumination which appeared dependent on the operating voltage of the ruby laser. When satisfactory, the schlieren lens was then introduced. This was located on the optical rail at 2f from the centre of the working section. Using a target located at the focal point of the lens, that indicated the position of the optical axis, the height and orientation of the lens was adjusted so that the focal spot was coincident with the axis. The imaging lens was then introduced and an object, usually a ruler to allow the system magnification to be determined, was placed in the working section. The imaging lens was then moved along the optical rail until an image of the ruler with the desired magnification was observed on the screen or in the camera view finder.

Having completed the optical alignment, the next stage of the set up procedure involved inserting the schlieren stop or knife edge into the system. As with the optical alignment, this is best done using a low intensity continuous source. An equally satisfactory result can be obtained using the pulsed ruby laser, provided the necessary eye protection is worn and the increased set-up time is no problem. The stop should ideally be positioned along the optical axis at the focal point of the schlieren lens. Incorrectly positioning the stop forward of, or immediately behind the focal point, usually results in a reduction in the system sensitivity. The most practical way of ensuring that the stop is located at the correct position is to observe the change in the image plane illumination as it is adjusted into the field. The basis of the method is shown in the schematic of figure 8, where rays converging to a focus at f are intercepted by a knife edge located at one of several points along the optical axis.

![Figure 8 Setting up the knife edge, after Foucault.](image)

At position 1, forward of the focal point, the observer would notice a region of shadow moving across the image plane from top to bottom as the knife edge is adjusted up into the field. A similar thing would be seen for a knife edge positioned beyond the focus.
at point 3, except here the shadow would now be observed moving in the opposite direction. In both cases, as the knife edge intercepts the optical axis half the field would be in shadow, as shown. Only when the knife edge is correctly positioned at the focal point of the system (point 2), will there be a uniform extinction of the image plain illumination. Theoretically total extinction should occur very rapidly within a short distance, since the width of the focal zone from the Rayleigh criterion is around 5 μm. In reality extinction is found to be more gradual due to the effects of diffraction and system aberrations on the width of the focal zone.

4.6.13 Synchronisation and timing

So far only the main components and their particular function within the system have been discussed. To get a fuller picture of how the images are actually generated, it is necessary to expand these discussions to include the subsystems that control the equipment. In addition to those parts already discussed, the system also includes several electronic delay devices, electrical and mechanical triggers, and sensing apparatus such as a photodiode and a 1000:1 voltage probe. The sensing devices are used purely to obtain timing information from the system, which is then displayed on an oscilloscope. In the case of the voltage probe, this monitors the discharge voltage across the EMAT coil. The photodiode on the other hand monitors the laser pulse as it emerges from the tank. If the system is correctly set up, the difference between the two signals should be equal to the time taken for the shock wave to enter the working section. A block diagram indicating the control structure from the initiating push button to the laser and EMAT is shown in figure 9.

![Figure 9 The control structure.](image)
The system has been designed to be initiated one of two ways, the first and main method is via the camera shutter contacts. In this mode the operator is required to depress the camera button, which starts the timing sequence and opens the camera shutter ready to take a picture. The time delay between the shutter opening, the arrival of the shock wave in the working section and its illumination by the laser is around 600 μs. The shutter must therefore be kept open for at least this length of time to be sure to capture the image. The upper limit, is dependent on the amount of parasitic light entering the system, in practice no problems were found with the standard shutter speed setting of 1/125 second. The second method of initiating the system is via a standard push button fitted with a debounce circuit. The main purpose of this method is to allow the operator to set and then check the system timing before taking a picture. In both cases a clean switching action is necessary for conversion to a TTL logic pulse, to prevent multiple pulsing and misfires. The signal is then passed onto the two delay devices, at the first, the signal is split with half being conditioned prior to initiating the laser firing sequence. The other half undergoes a pre-set time delay followed by a further variable delay before initiating the spark gap breakdown sequence and firing the EMAT. The total length of the delay which includes the system and shock propagation times is calculated so that the shock wave coincides with the laser pulse initiated 600 μs earlier. The complete timing sequence for the system is shown in figure 10.

![System timing sequence diagram](image)

**Figure 10** The system timing sequence.

Within 35 ns of the system being triggered by the shutter contacts on the camera, the laser begins its comparatively long but very stable firing sequence. There then follows a pre-set delay of 500 μs and a user defined delay of anything between 1-100 μs (to take account of the shock propagation time) before the preionising spark is applied to the spark gap. At this point we must wait for a suitably placed electron between the
electrodes to begin the avalanche process, that marks the electrical breakdown of the gap and the flow of current. Reference to section 3.311 will show that this is a very complicated process. The reader may also be aware that factors such as electrode spacing, charge voltage and the state of the gas (in this case air) will effect the discharge and in particular the delay. Other factors producing changes during operation will lead to additional intermittent delays.

Even with a correctly set up and well maintained spark gap, intermittent delays of between 1-100 μs were found to occur. As such this was the main cause of misfire events in the system, where the shock wave was either to early or arrived to late to be illuminated by the laser. The standard spark gap delay was usually no more than 2 μs, slight variations around this figure, knowing that the shock propagated at a speed of approximately 1.5 mm/μs, were generally acceptable. Once conduction across the gap is initiated, the discharge lasts for approximately 6 μs, during which time a strong acoustic transient is emitted by the EMAT. Assuming all other delays in the system remain constant, the point at which the shock wave is photographed is determined by the variable delay set by the operator. Any setting between 1-100 μs will ensure capture of the shock wave within the boundaries of the working section. The smaller the delay the more advanced the position of the shock wave in the working section.

4.62 The precision X-Y translation stage

The tank mounted positioning system shown in figure 11, was initially constructed to provide a stable platform for hydrophone measurements of the acoustic field. The basic feature of the system were the two steel bars or runners that extend along the length of the tank, (x-axis) which provide the support track for a computerised positioning translator and the acoustic lens support and positioning apparatus. The computer controlled translation stage, consisted of individually motorised horizontal (z-axis) and vertical (y-axis) translation slides, allowing the hydrophone to be positioned any where on the y-z plane. Each motor was individually connected to a purpose built stepper motor controller, which was interfaced to an IBM compatible personal computer running a stepper motor control program. Transducer positioning along the x-axis was controlled manually, by physically pushing the translation stage along the guiding rails and then locking it into its new position. Manual control of both translation slides was also possible, via the thumb wheels at the end of each lead screw. A multifunctional tool head connected to the vertical slide allowed further transducer manoeuvrability, by enabling it to be rotated around the y-axis or tilted
about the z-axis. The positioning system was also used to support simulation kidney stones during shock interaction studies and for the support of grids and other structures in scale method measurements of the acoustic field.

Figure 11. The tank mounted positioning system.

4.621 The computer interface

The interfacing between the IBM compatible computer and the stepper motor driver was facilitated using a commercially available data acquisition and control card supplied by Eltime Ltd (figure 12). The half size card was fitted into one of the computers 8 bit 62 pin I/O expansion slots. As well as allowing read and write operations on the data bus the card also permits direct memory access (DMA) transfers. The cards inputs consisted of eight data lines (SD0-SD7), a base (SA4-SA9) and offset address (SA0-SA3), the I/O read/write lines -IOR and -IOW and the address enable (AEN) and address latch enable (Bale) bus signals. From the output of the card a 25 way ribbon cable transfers the control information to the stepper motor driver hardware.

Before any step information can be transferred along the data bus to the stepper motor controller, a valid base address has to be present on the address bus (SA4-SA9). When correct, the bale line is taken low latching the address onto the D-type flip flop (74LS374). This is then compared by the 74LS682, an 8-bit comparitor, with the preset address on the DIL switches. Which in this case was set to 300H. If the address on the address bus is correct, the output of the 74LS682 is taken low, resulting in one of the Nor gate inputs (74LS54) also going low. Removing the remaining blocks on the
74LS54 and thus enabling the octal tri-state transceiver (74LS245) which buffers the data bus, requires the AEN and either the -IOW or -IOR lines also to be pulled low. Under normal circumstances AEN will be low, and will only become active if an external device requests a DMA transfer. Therefore, by setting either the read or write lines low, the output of the 74LS54 goes high, enabling the transceiver and placing the clock pulse \( cp \) on the output.

Figure 12 The Eltime interface card.

The direction of the data flow is controlled by -IOR, when high, data is transferred from the computer bus to the output. The data's destination address is determined by the first 4 bits \( (SA_0-SA_3) \) of the address code and must be added to the base address of 768\(_{10} \) (300H). Hence to access the stepper motor function associated with the address \( A_0 \) we require the binary equivalent of the base address 768\(_{10} \) plus the address of \( A_0 \) (1) giving 769\(_{10} \) to be placed on the address bus. Every stepper motor function has a particular address which can be accessed by first placing the correct address on the bus before transferring the relevant control data. When the data transfer is complete -IOR is taken high, disabling the transceiver and removing the clock pulse from the output preventing further transmission.

4.622 Stepper motor hardware

The stepper motor controller shown in figure 13 was designed around two SAA1027 stepper motor drivers, each of which is capable of controlling one 4 phase bi-
directional stepper motor. Fitted to each of the motors was a synchronous gearbox which reduced the step angle from the standard 7.5° to 0.15°. This 50:1 reduction was necessary to increase the motors output torque, which without the gearbox was too low to drive the translation stage. The use of the two dedicated driver IC’s meant that the only other circuit components required were those associated with the decoder logic. Here, the binary code on the address and data bus is first decoded and then converted into stepper motor instructions according to the specific address and data map for the system. Table 1 shows the address map for the stepper motor controller, the address of the required function is shown enclosed by the brackets [ ] and listed immediately below is the data required to drive the stepper motors.

<table>
<thead>
<tr>
<th>Device</th>
<th>Direction [769]</th>
<th>Step [770]</th>
<th>Axis</th>
</tr>
</thead>
<tbody>
<tr>
<td>D#1</td>
<td>4 (Left)</td>
<td>4</td>
<td>Z</td>
</tr>
<tr>
<td>D#1</td>
<td>8 (Right)</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>D#2</td>
<td>2 (Up)</td>
<td>1</td>
<td>Y</td>
</tr>
<tr>
<td>D#2</td>
<td>0 (Down)</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 The stepper motor controller address and data map.

An example of a command to move the translation stage to the right along the z axis would consist of the direction command with the address [769] and the data 8 [10] (for clockwise rotation), followed by the step command [770] and the data 4 [10] (device 1). The input requirements of the driver IC’s meant that the direction command must always proceed the step request.

On presentation of the direction command at the input of the controller, the data component of the instruction appears at pins 2&12 of the 74LS74 D-type flip flop. Because the data bus is tri-state and tends to float at around 2 Volts when not in use, a delay of 50 ns (facilitated by the three And gates in series) was found to be necessary, before the data became valid and the address component of the instruction could clock the data to the output of the 74LS74 via pins 3&11. The direction signal is then converted by the HEF4104BP low to high voltage translator, from TTL to the high voltage CMOS logic level required by the stepper motor drivers. Once the direction has been set the 'step' command is sent, the data is clocked onto the voltage translator by the clock and the motor stepped 7.5° in the desired direction. The direction-step procedure is then repeated for each of the subsequent steps of the stepper motor. By noting both the gearbox ratio and the thread pitch of the slide lead screw, the number of steps required to move the slide a set distance was calculated. This conversion factor was then incorporated into the software to allow the user to input distances in mm’s rather than defining the number of steps.
Figure 13 The stepper motor controller.
4.623 Control software

The computer program used to control the y-z translation stage was written in Turbo Pascal. It was based on the Pascal command Port[Address]:=Data, which instructs the computer to send along the I/O bus the data D to the device at the address [. A key feature of the program is the choice of two scanning modes, one of which the user is initially prompted to select. The simplest, is the move operation in which the tool head can be moved to any position in the y-z plane, subject to the maximum travel. On completion of the move operation, the user is then returned to the original screen prompt. At this point the user may then wish to select scan mode, in this mode the user can define an area on the y-z plane of y mm by z mm in which they are particularly interested. A further prompt then allows them to select a scan interval (in mm's), where the tool head is held stationary while a measurement or an operation under manual or computer control is undertaken. The tool head is then moved on a distance equal to the scan interval until a sweep, z mm long has been completed. The head is then moved down and the operation repeated until a series of such horizontal sweeps is built up and the defined area has been covered. When complete, the tool head is then returned to its start position. The resolution and translation speed of the system is determined to a large extent by the stepper motor gearbox ratio. The earlier motor torque problems seen with certain ratio's, required an extra degree of program flexibility in the form of user defined gearbox ratio. This allowed different gearbox ratios to be used with the stepper motors without the need to rewrite and compile the program. The design and development of the positioning system and its interfacing with the computer was found to be a very useful and instructive exercise. The system has found extensive use for measurements and equipment positioning operations throughout the period of research.

References

3 Schardin, H. Schlieren methods and their applications. RAE translation 122.


CHAPTER 5

SCHLIEREN STUDIES

5.1 Experimental method

The generation of shock waves and their observation using the various schlieren techniques formed the two main areas of interest during the period of research. The first, which was essentially concerned with building and then testing the EMAT, hydrophone and optical apparatus etc., has been described in detail in the previous chapters. In the remaining part of this thesis we shall therefore concentrate on the second and possibly the most interesting phase of the work, namely the schlieren and shadow studies of EMAT shock waves. The results of these studies, from which a major part of the seven publications (appendix IV) that accompany this work have been derived, will be presented in this chapter.

As an introduction to the work, the motivation and methodology behind the various second phase studies will now be investigated. In addition to the characterisation and quantitative analysis of the schlieren images generated by the shock waves, studies into cavitation and the effect of shock waves on simulation kidney stones have been carried out. The priority was of course understanding the schlieren and shadow images of the planar and focused shock waves. The initial work concentrated on how the structure of the shock wave changed during its passage through the working section. Of particular interest was its shape at the focus, in the centre of the working section.

For this study, the system was set up so that the shock wave was initially captured as it emerged from the acoustic lens, into the working section (maximum delay). For each of the subsequent shots the EMAT delay was steadily reduced, until, a sequence of photographs of individual planar and focused shock waves propagating through the working section was built up. As the individual images represented separate events, it was important that the discharge conditions remained constant. This was essential if the later comparisons between the images were to have any validity. One of the most significant problems was determining the relationship between the temporal profile of the shock wave and the multitude of components in the schlieren images. To help analysis, the opaque spot spatial filter was replaced by a knife edge. By positioning the knife edge to select either positive or negative values of the density gradient, the
components of the image could be directly related to that particular part of the shock wave.

A natural progression of the work with the knife edge spatial filter was its use to obtain quantitative information (section 4.4) about the shock wave. As previously mentioned, pressure variations within the acoustic field can be directly related to the angular deflection of rays in the Fourier plane of the schlieren system. Quantitative information from the system can therefore be obtained by noting the relative displacement of the knife edge necessary to eliminate rays associated with a particular component of the image. By repeating this for differing levels of filtering, a comprehensive assessment of deflection angle and thus pressure associated with each part of the schlieren image was made. An alternative way of obtaining quantitative information from the system was to use the scale method. This required a grid or a system of lines to be positioned in the field just behind the shock wave. After replacing the knife edge with the opaque spot, the schlieren system was refocused so that the grid and not the shock wave sat in the object plane of the system. A quantitative assessment of a section through a planar shock wave was then made from the observed displacement of the lines as the planar wave passed through the working section. The results obtained from the schlieren studies were then compared with single point hydrophone pressure measurements of the acoustic field.

Of those areas in the field of lithotripsy that still remain only partially understood, it is the mechanisms of shock wave induced stone disintegration that are by far the most important. Despite much research, the contribution of the individual mechanisms is still not clear. The reason for the lack of progress, is primarily because of the difficulties of observing the actual interaction in vivo. Those studies that have attempted to observe the interaction under simulated conditions, have usually done so with limited success. In many cases this can be attributed to the method of observation. The application of laser illuminated schlieren methods to this problem, has for the first time given us the opportunity to look in far greater detail at the interaction process. In addition to cavitation, which is physically visible, albeit for a very short time, it is possible to see the actual shock wave and the cavitation shock transients. In the first of the interaction studies to be presented, shock waves were incident on a circular brass target located in the focal region. The actual position of the target was adjusted, to either fully or partially obstruct the shock wave during its passage through the working section. The images of the interaction, particularly those of the partially obstructed shock wave have yielded some important details. However, the real benefit, is seen in a sequence of images showing the EMAT shock wave
during its interaction with a simulation kidney stone. The incident, transmitted and reflected components of the shock wave are all clearly visible, along with cavitation and even stone debris. The damage sustained by the stone was also monitored, most notably after only 40 shots, to allow a direct comparison to be made with another study. Studies of 200 shots, where both entrance and exit stone damage was sustained will also be presented.

Cavitation is suspected of playing a major role in the disintegration of stones during lithotripsy (section 2.541). The formation of the gas or vapour filled bubbles, occurs in response to the reduction in pressure and rupture of the water by the rarefaction phase of the shock wave. The cavities so produced are stimulated into motion by the remainder of the shock wave, and either oscillate non linearly about some equilibrium point or expand to many times their original size. Cavities that expand in this way are relatively short lived and often collapse violently, emitting very strong spherical shock transients. The production of cavitation shock transients was normally seen to occur only in the tail and the field immediately behind the shock wave. One interesting finding of the schlieren studies, has been the non symmetrical appearance of the normally spherical cavities and cavitation transients when enveloped in the shock wave. The reason behind this asymmetry has therefore been investigated. It is proposed that the distortion is associated with an optical effect introduced by the compression and rarefaction regions of the EMAT shock wave. The study and analysis of this lensing effect has been carried out using a technique based on the scale method. To compare the effect on the transients in the higher pressure regions of the shock wave, the field was seeded with micro bubbles produced by circulating water through an impeller type water pump. This led to the generation of a few cavitation events immediately behind the main shockfront, and therefore allowed a direct comparison of the cavitation transients occurring in both the high and low pressure regions of the EMAT shock wave. The cavitation transients were then modelled by placing a perspex sheet (PMMA) with a set of concentric circles scribed on it, in the object plane of the schlieren system, a known distance behind the focus of the EMAT. The system of circles was later replaced by a system of lines or grids, from which the contribution of each part of the shock wave field to the observed distortion could be seen.

5.2 Schlieren and shadow studies

The transducer studies of the acoustic field generated by the EMAT (chapter 3) have played an important role in understanding some of the properties of these devices.
Basic hydrophone measurements have shown that the shock waves, which comprise of a number of compression and rarefaction cycles, are cylindrically symmetric about the EMAT axis. This has been found for both the planar and focused shock waves, although in the case of the focused wave the amplitude was found to decrease rapidly with radial distance. Along the length of the shock wave, the amplitude of the compression and rarefaction phases of the shock wave were found to decrease towards the tail, following the discharge current. Despite the obvious success of the hydrophone studies in characterising the shock waves, little could be found about the other effects, such as cavitation, known to be present in the fields of these devices. In an attempt to observe this phenomenon and see how it might contribute to the stone fragmentation process, the shock waves were observed using various schlieren techniques. The schlieren and shadow images produced by the shock waves have yielded some important details about the shock wave and cavitation in the field. However, the complexity of the images and the limitations of the system have made analysis difficult. Fortunately, by combining the information from the hydrophone studies with that obtained from the schlieren studies, it has been possible to analyse most of the effects observed in the images.

In the first of the studies, an opaque spot spatial filter was used to obtain schlieren images of planar and focused shock waves. The schlieren and shadow images of a planar shock wave taken 92 μs after firing the EMAT are shown in figure 1a,b. The images can be visualised as disks of water of varying density, viewed edge on, propagating from left to right at approximately the speed of sound in water, (1480 ms⁻¹) in a direction parallel to the plane of the page. A more appropriate description in the case of the schlieren image (figure 1a), is that the bands or components actually represent the density or refractive index gradients of the shock wave. Within each of the five bands, in the more sensitive schlieren image, interference type fringes are visible. These have been brought about by light that has been diffracted in the system and superimposed upon the normal Fraunhofer pattern in the focal plane of the schlieren lens. A similar phenomena was reported by Gayhart and Prescott² in a study in which they investigated these schlieren fringe patterns as a means of obtaining quantitative information about the inhomogeneity. On the leading edge of the first of these components i.e. the first on the right, the shock front can be observed. The very fast rise time and large density gradient of the shockfront has produced what is essentially a narrow bright line. In the less sensitive shadow image of a planar wave (figure 1b) only the first two components are visible, however, the shock front can once again be seen at the front of the wave preceding the two shadow regions. The total length of the disturbance from the schlieren image is approximately 15 mm, at
Figure 1a The schlieren image of a planar shock wave. Scale bar = 10 mm.

Figure 1b The shadow image of a planar shock wave.
1.48 mm μs⁻¹ this makes the total disturbance 10 μs long, which is in good agreement with the earlier hydrophone measurements. Similar measurements of the shockfront indicate a rise time of 300 ns. Immediately behind the shock wave in both images cavitation has taken place, this is visible as the dark spots in the shadow image or the small bright ring like structures of the schlieren image. In both cases the spatial distribution of the cavitation is restricted to the central region, confirming the Gaussian-like pressure profile observed in the hydrophone study of the planar shock wave. The additional structures present in the field of the shadow image are interference and diffraction phenomena introduced by the system components (regular concentric circles) and by material in the water.

The inability of most planar sources to generate the high energy densities required for lithotripsy, usually means that a focusing device of some kind or other is required. However, focusing devices, such as lenses, by their very function lead to changes in the shape and characteristics of a shock wave. As we also utilise a lens, it was important to see what additional effects might be present in the focal region of the system. Figures 2-4 show a sequence of schlieren and shadow images of focused shock waves taken 62, 86 and 92 μs after firing the EMAT. In the images this can be seen to correspond to a shock wave positioned at the focus of the system and at the two extremes of the working section. In the first of the sequence figure 2a,b, the shock wave has been imaged emerging from the acoustic lens. A comparison between these images and those of the planar images (figure 1a,b) indicates several recognisable similarities, albeit in a wave that is saucer shaped as it converges to a focus. However, if we neglect the converging form of the schlieren and shadow images of figures 2a,b, a similar form and number of bands in each of the two images can be observed. The main structural difference between the planar and converging shock wave is the position of the shock front. In the planar wave this was observed leading the components of the two images, but in figure 2a,b the shock front or bright line can be seen in the middle of the first band. This is particularly clear in the shadow image of figure 2b. The reason behind this apparent shift in position is due to the saucer shape of the shock front, rays passing through the periphery of the saucer will be affected less than those passing through the central region, tangential to the shock front. The end result is an image of a saucer shaped shock front with a strong base but weak and blurred sides.

At the focus (figure 3a,b) the shape and structure of the shock wave has changed considerably. The higher pressures/gradients in this part of the field has resulted in much stronger effects being produced in the images. The most notable is due to the
shockfront but other effects, such as the fifth component of the shock wave, previously only visible in the schlieren images is now visible in the less sensitive shadow image. One of the most significant findings is the presence of a second but weaker shock front in the third component of the shock waves. The formation of this second shockfront is no accident, since the same physical effects that have produced the main shock front are also acting on the other parts of the wave. The third component is in fact the second compression phase of the shock wave, which due to non-linear distortion has also developed a shock like structure. The relationship between the other components and the temporal profile of the shock wave will be highlighted shortly. After passing through the focal region, the shock wave begins to diverge resulting in a reduction in the intensity of the wave. However, cavitation events are still being generated resulting in a trail of these events extending the whole length of the working section. This is seen in the third set of images (figure 4a,b) where the shock wave is seen just about to exit the working section on the right.

Cavitation activity is once again visible, particularly in figures 3 and 4. The increased activity has been brought about by the rupture of the water by the very much stronger negative pressure region of the focused shock wave. Tensile strengths for water as low as -0.5 MPa \(^3\) and up to -20 MPa \(^5\) have been reported, amongst other things its strength appears to be dependent on the gas and impurity content of the water. The cavitation bubbles seen in this field indicate that the acoustic cavitation threshold or tensile strength of the water was below -6 MPa i.e. that exerted by the main rarefaction component of the shock wave. This suggests that either impurities in which gas pockets have been stabilised against dissolution, or gas in the form of pre-existing micro bubbles from the preceding shock excitation of the medium were present in the water during the experiments. The cavitation bubbles first appear after the second component in the schlieren image, suggesting that this is the location of the rarefaction phase of the shock wave. Proof of this was obtained from the subsequent knife edge studies of a focused shock wave figure 5a,b. The results indicate that the rising edges (negative values of \(dn/dx\) in our co-ordinate system) of the shock wave i.e. the compression phases, are represented in the image by the first, third and fifth components (figure 5a). The falling edges or rarefaction phase of the shock wave (positive values of \(dn/dx\)) being represented by the second and fourth components (figure 5b). The reader may prove this for his/herself by mentally superimposing one image on top of the other, and then comparing it with the image of the focused shock wave produced using an opaque spot spatial filter in figure 3a. The components in the schlieren images of figures 1-4 are therefore, alternately from right to left, negative and positive values of the refractive index gradient i.e. the rising and falling edges of
the shock wave. A direct comparison between the fourth and fifth components in the images with the temporal profile of the shock wave generated by the hydrophone (section 3.64) was not possible. This was due to the limitation of the hydrophone to accurately reproduce the tail section of the shock wave.

Using the information gained from the knife edge study it is possible to see that the cavitation bubbles first appear after the first negative phase of the shock wave on the rising edge of the second compression (3rd component in the schlieren image). They continue to expand and oscillate as they experience further compression and rarefaction cycles, until finally collapsing and emitting cavitation shock transients on the rising edge of the third compression, (between the fourth and fifth components) or remaining for a time as stable bubbles in the field. The cavitation shock transients can be seen in both sets of schlieren and shadow images as the bright and dark rings respectively. As these 'primary' cavitation transients expand behind the main shock wave, they appear to interact with some of the remaining bubbles in the field producing 'secondary' cavitation collapse. This is seen in both the schlieren and shadow images as the mass of cavitation events in the shock wave wake to the left of figures 3 and 4. It has been suggested\(^6\),\(^7\) that the processes involved in producing these 'secondary' cavitation events may be very different to those producing the 'primary' events. The main difference is the pulse duration of the shock front. For the main shock wave, our studies have indicated values of between 100-500 ns. However, studies conducted by Vogel and Lauterborn\(^8\) indicate that the duration of a shock transient emitted on bubble collapse is around 30 ns (half width maximum) i.e. approximately 50\(\mu\)m. Consequently, even small bubbles are not completely enclosed within the pressure region\(^6\), but rather a very narrow pressure wave passes the bubble and accelerates only a small part of the bubble wall at any one instant. The remainder of the bubble wall rests in pressure equilibrium with the normal static pressure. Our schlieren studies certainly seem to indicate a difference between the 'primary' and 'secondary' transient. Careful examination of the 'secondary' transients in the images show that these transients have a double shock structure, the separation being approximately 300 ns. Whether this has occurred as a result of excitation by the 'primary' shock transients is not known, further study is therefore required. In addition to the shock transients, a significant number of bubbles can be seen in the field, some of which are quite large (\(\approx\)0.71 mm in diameter). Bubbles this size would normally require a large amount of inward gas diffusion (rectified diffusion\(^9\)) and several thousand acoustic cycles. A possible explanation given by Coleman et al\(^{10}\) is that these bubbles result from the expansion of some nuclei to such relatively large volumes, that they encompass significant amounts of gas during evaporation of the liquid that occurs
Figure 2a The schlieren image of a converging shock wave near the lens. Scale bar = 10 mm.

Figure 2b The shadow image of a similarly positioned shock wave.
Figure 3a The schlieren image of the converging shock wave at the focus.

Figure 3b The shadow image of a similar shock wave.
Figure 4a The schlieren image of a diverging shock wave.

Figure 4b The shadow image of a similar shock wave.
Figure 5a The focused shock wave components associated with negative values of \( \frac{dn}{dx} \).

Figure 5b The focused shock wave components associated with positive values of \( \frac{dn}{dx} \).
during bubble expansion. If enough gas is collected, and the collapse is not sufficiently violent to shatter the bubble, then some bubbles containing enough gas to be visible could result.

The forward propagating component of the 'primary' cavitation transients can be seen to bunch up in the region associated with the original collapse. This occurs because the individual transients, although being formed at different times, propagate through the water at the speed of sound (1480 ms⁻¹) and thus keep up with the main shock wave. Analysis of the images show that these transients form an additional shock region i.e. one composed entirely of cavitation shock transients.

5.3 Quantitative studies

5.31 The assessment of pressure using a knife edge

The images presented so far have highlighted some of the benefits of using optical techniques to analyse acoustic shock phenomena. Reference to chapter 4 shows that in certain circumstances it is also possible to obtain quantitative information from these images. In the first of two studies to be presented on the quantitative analysis of the shock waves generated by an EMAT, a knife edge will be used. Similar studies have used opaque disks of varying diameters, however, while there are certain advantages to using this technique (chapter 4.32) the author feels that in practice it is a little cumbersome. In the second quantitative study the scale method will be used. The use of a knife edge spatial filter in the manor described earlier in this chapter, (section 5.1) has resulted in the production of a sequence of planar and focused schlieren images, figures 6a-d and 7a-d respectively. For this particular study only those components associated with negative values of dn/dx have been selected, in order to observe the rising edges of the shock wave. This can be confirmed in the focused case by comparing figure 7a with the knife edge study of a focused shock wave in figure 5a (appendix IV).

Analysis of the images in figures 6 and 7 show that as the level of spatial filtering is increased, so the components associated with the weakest parts of the shock wave are removed from the image. This is particularly apparent for those components produced by the tail of the shock wave where the pressure is lowest, although, rays that have passed through the periphery of the saucer shaped focused shockfront are also rapidly lost. Cavitation transients (figure 7a) are similarly obscured by low levels of spatial
Figure 6a The quantitative study of the components associated with negative values of $dn/dx$ in a planar shock wave. Components in the image producing total deflections equivalent to pressures of $\geq 162 \text{ kPa}$.

Figure 6b Components in the image producing total deflections equivalent to pressures of $\geq 0.59 \text{ MPa}$.

Figure 6c Components in the image producing total deflections equivalent to pressures of $\geq 1.66 \text{ MPa}$.

Figure 6d The component associated with the main shock front $\geq 2.4 \text{ MPa}$.
Figure 7a The quantitative study of the components associated with negative values of $dn/dx$ in a focused shock wave. Components in the image producing total deflections equivalent to pressures of $\geq 380 \text{ kPa}$.

Figure 7b Components in the image producing total deflections equivalent to pressures of $\geq 2 \text{ MPa}$.

Figure 7c Components in the image producing total deflections equivalent to pressures of $\geq 5.7 \text{ MPa}$.

Figure 7d The component associated with the main shock front $\geq 19.2 \text{ MPa}$.
filtering, despite having in some cases pressures as high as 5.8 Mbar\textsuperscript{13}. The last two cases in particular highlight one of the limitations of obtaining quantitative information using schlieren methods. This is that ray displacement is not only dependent on the density gradient but also on the extent of the inhomogeneity c.f. Eq. 4.13. The final images in the sequences (figures 6d and 7d) show the component associated with the shockfront itself. As expected, this part of the wavefront produces the largest ray displacement, in the case of the planar shockfront this was measured to be $14.25 \times 10^{-3}$ m. A similar measurement made for the focused shockfront yielded a ray displacement of $18 \times 10^{-3}$ m in the focal plane of the system.

Using these figures and those obtained for the other components observed in the two fields, a quantitative assessment of shock wave pressure was made. However, reference to equation 4.17, tells us that in order to do this it is also necessary to know the extent of the disturbance $\Delta z$ through which the ray has passed and the average rise distance $\Delta x$ (sound speed x rise time) in the field being measured. For this study $\Delta z$ was assumed to be the distance over which the shock wave can be calculated, still to produce a positive contribution to ray deflection, i.e. we can neglect the small contribution of the lower pressure regions found at the periphery of the shock wave. For ease this was also the distance over which $\Delta x$ was assumed to have a constant value. The actual value of $\Delta x$ for the planar and focused shock waves were obtained from the hydrophone measurements of the rise time (section 3.64) in both fields. In the planar case a figure of 500 ns was selected, while for the focused shock wave at the focal point 100 ns was used. The value of $\Delta x$ for the planar and focused shock waves was therefore calculated to be $7.5 \times 10^{-4}$ m and $1.49 \times 10^{-4}$ m respectively. Evaluating equation 4.17 for the planar shockfront (figure 6d), for $\Delta z=60 \times 10^{-3}$ m and $\Delta x=7.5 \times 10^{-4}$ m ($1480$ ms$^{-1} \times 500 \times 10^{-9}$s) gives a pressure of around 3.2 MPa. Analysis of the focused shockfront (figure 7d) is a little more complicated due to its saucer or semi hemispherical shape, however, at the focal point $\Delta z=2.5 \times 10^{-3}$ m and $\Delta x=1.49 \times 10^{-4}$ m ($1480$ ms$^{-1} \times 100 \times 10^{-9}$s), giving rise to a pressure of 19.2 MPa.

5.32 The assessment of pressure using the scale method

Building up a quantitative picture of a shock wave using the knife edge method required the production of several unrelated shock wave images, each of which required a different level of spatial filtering. Apart from concerns about the reproducibility of the shock waves produced by the EMAT, the process tended to be very time consuming. One of the great strengths of the scale method (section 4.32) is
that a quantitative assessment of most of the shock wave can be obtained from a single image, thereby improving the time and consistency of the results. Apart from the grid or system of lines positioned in the object plane, the set-up is essentially that of a conventional schlieren system. When the grid is viewed through an optical inhomogeneity such as a shock wave, lines on the grid appear displaced. Figure 8a-d shows the shadow images of a planar and focused shock wave, passing in front of and displacing a system of lines. The actual quantitative measurements were made on a section through the planar wave as indicated in figure 8a, using a system of lines with a spacing of \( \approx 1 \times 10^{-3} \) m. This particular characteristic of the grid is not critical, and so can vary to quite a large extent without affecting the accuracy of the measurements. What is important is that the grid lines are straight and parallel to each other. Careful examination of this image will show that the lines beyond the immediate vicinity of the shock wave remain undistorted, while those within are displaced to the left or right, by the rising and falling edges of the shock wave. By measuring the amount by which these lines are displaced to the left or right of the original intersect with the section through the shock wave, the pressure, positive or negative, can be calculated. From this data a partial reconstruction of a planar shock wave has been attempted. As with the knife edge study, the ray displacement data alone is not sufficient to calculate the pressure, once again we need to know the extent of the field \( \Delta z \) and the rise distance \( \Delta x \). Thus as before the hydrophone figure of 500 ns for the rise time will be selected and is assumed to be constant over the distance \( \Delta z \), across all three planar wavefronts. Also \( \Delta z \approx 60 \times 10^{-3} \) m for the first wavefront and around \( 40 \times 10^{-3} \) for the second and third wavefronts.

Rearranging equation 4.12 section 4.32 for \( n \), noting that \( \frac{dn}{dp} = 1.5 \times 10^4 \), where \( dp \) is the change in pressure, and given that the system magnification \( M \) in the original working image was 3.15 and \( b = 63 \times 10^{-3} \) m, the pressure associated with each of the measured displacements was calculated. For reasons of compactness the system magnification factor was reduced to 2 for the schlieren image of the planar wave (figure 8a) contained in the thesis. By plotting the calculated pressure against the temporal position of each of the intersect points on the section through the planar shock wave, the wave was reconstructed (figure 9). Analysis of figure 9 clearly shows the three compression phases of the focused shock wave, previously unresolved in the hydrophone measurements. This underlines one of the benefits of using a high resolution, non invasive optical technique to study acoustic shock waves. An important point to note here is that in the region associated with the rising edges of the shock wave, particularly the shock front with its large refractive index gradient, the full displacement of certain grid lines cannot be resolved. In these cases the grid is
Chapter 5: Schlieren studies

Figure 8a The shadow image of a planar wave displacing a system of lines. Width of the scale bar is 10 mm.

Figure 8b The shadow image of a focused shock wave displacing the same system of lines. Width of the scale bar is 10 mm.
Figure 8c The focused shock wave displacing a system of lines at the focus.

Figure 8d The focused shock wave displacing a system of lines at the edge of the field.
bent and displaced to such an extent that the image of the grid is lost to the system or too weak to be imaged. Hence the displacement data is missing for these particular parts of the graph. This appears to be one of the drawbacks of using this technique to obtain quantitative information from a schlieren image. This is also the main reason why no quantitative analysis was undertaken on the focused shock wave, where large density gradients exist throughout the wave.

![Graph of Pressure vs Time](image)

Figure 9 The reconstructed temporal profile of a section through a planar wave, produced from the line displacement data of figure 8a.

### 5.33 Discussion

The use of the knife edge and scale methods to obtain quantitative information from schlieren images generated by the EMAT shock waves, have benefited the work in two key areas. The first and possibly the most useful, has been the identification of the components within the schlieren and shadow images. This has made it possible to relate specific components to their transducer counterparts, therefore improving the understanding of these complex images. It has also been possible to make a quantitative assessment of shock wave pressure. However, often only an estimate of the pressure is possible, since with a complex inhomogeneity, the angular deflection will be a result of the summation of all the deflections due to $\frac{dn}{dx} \Delta z_i$. Although the
extent of the inhomogeneity in the z direction is determined relatively easily, the value of dn/dx at each point in the field is not. For example, rays travelling through the central region of the axially symmetric wavefront will undergo the greatest deflection, since they not only encounter the highest pressure gradients but also traverse the whole width, z, of the field. Rays traversing regions above and below the central region may encounter equally high pressures but pass through less of the wavefront, resulting in less deflection. The quantitative images will only show the degree to which a ray has been deflected. They are not sensitive to how the deflection was produced. Therefore, quantitative analysis of a shock wave using either of the methods cannot accurately determine the pressure, unless hydrophone measurements have first determined the value of dx for every value of Δz. The calculation of dn/dx and Δz for a complex inhomogeneity underlines the difficulties of assessing quantitative data from the schlieren image without having prior information about the inhomogeneity itself.

The result for the quantitative analysis of the planar field using the knife edge method has indicated a peak pressure of 3.2 MPa. Single point hydrophone measurements of the same field show the peak value to be nearer 4.6 MPa; the schlieren result would therefore appear to be in error by in excess of 50%. It is suggested that the value obtained from the knife edge study is not a peak value but reflects more the average pressure for that part of the shock wave. This is not totally unexpected, as the assumption that the rise time and thus Δx is constant across the wavefront (Δz), brought about because of the difficulties described above, has an averaging effect on the results. Since the same assumptions were also made for the quantitative study of a planar wave using the scale method, it is fair to assume that similar errors are likely to exist. The results obtained for this study indicated a pressure for the shockfront somewhere around 4 MPa. A more precise figure for the pressure cannot be obtained due to the difficulty of observing and measuring the ray displacement associated with this part of the shock wave. The reconstruction of the whole shock wave (figure 9) from the line displacement data of figure 8a shows the two additional compression cycles that follow the main shockfront. Whilst the data points are temporally correct and indicate the correct positions of the compression and rarefaction phases of the shock wave, the failure of the hydrophone measurements to resolve this area of the field, means that the values of Δx used in the calculations cannot be confirmed. As a result the pressure assigned to this part of the shock wave may be inaccurate. This inaccuracy may be further compounded by the errors introduced when measuring the ray or line displacements, and may go some way to explaining the difference between the pressures obtained with the two schlieren methods.
The complex nature of the focused shock waves, which are characterised by rapidly changing highly peaked pressures and rise times across their wavefronts, precludes the use of a constant $\Delta x$ over a fixed path length. The relative accuracy of the quantitative knife edge study of a focused shock wave (figure 7d) was due in part to the small dimensions of the focal region and relatively accurate hydrophone measurements of the rise time. However, complications arise when one considers the wave as a whole, since as well as the characteristics mentioned above, the components are saucer shaped. These and the problems associated with resolving the displacement produced by even the mildly shocked planar wavefronts, were the main reason for not analysing the scale images of the focused shock waves in figures 8b-d. The conventional method for extracting radial information in circularly symmetric systems is to apply the Abel inversion technique\textsuperscript{14}. In principle this technique could be used in conjunction with the scale images, here the distortion at large radii would be used to calculate the refractive index gradient and data from paths nearer and nearer to the axis. Thus progressively building up a radial index profile, especially of the lower pressure regions of the shock wave. Whether the technique is actually applied to the analysis of the schlieren images will depend on the direction of future research and on the practical benefit of such analysis.

5.4 Shock wave interaction studies

5.41 Interaction with a brass disk

So far, the schlieren studies presented in this chapter have only been concerned with the analysis of the actual shock waves in the schlieren and shadow images. With those discussions in mind, it is right that we now move on to examine some of the practical benefits of observing shock waves with schlieren techniques. One of the prime objectives of these practical studies was to observe a shock wave during its interaction with a kidney stone. Thus helping to confirm the presence of stone fragmentation mechanisms such as cavitation, liquid jets and spallation, all of which have attracted considerable interest in this field of research (chapter 2.54). However, before we consider that particular study we will look at a similar interaction, one in which a focused shock wave is incident upon a circular brass disk. The disk was positioned at the focus of the acoustic lens, but off axis, such that half the shock wave was obstructed during its passage through the focal region. Figures 10 a,b show the shock wave during and after its interaction with the brass disk. Prior to the interaction, the shock wave looked very much the same as that recorded in the schlieren image of
Figure 10a The schlieren image of a shock wave during its interaction with the brass disk. Scale bar = 10 mm.

Figure 10b The shock wave after its interaction with the brass disk.
Figure 3a. Close inspection of figure 10a will show that the obstructed part of the shock wave has, as expected, been partially transmitted and partially reflected. The transmitted part, which at this point only consists of the shock front, has after being reflected at both interfaces, emerged from the brass disk. The attenuated shockfront 'A' can be seen to have shed its surrounding detail and be physically displaced from that of the unobstructed part of the wavefront. The components that previously surrounded the main shockfront have been attenuated and are now too weak to be imaged. Its advanced position is due to the higher acoustic velocity in brass (4373 ms\(^{-1}\)). The physical displacement of the shockfront with respect to the unobstructed shock wave, allows us to calculate and thus confirm the original location of the shockfront 'B' as being the same as that seen in figure 2a,b.

The reflected component is similarly visible, its original form still discernible amongst the very active cavitation field the initial reflection has produced. Leading this part of the wave back towards the lens is the reflected component of the shockfront, part of which appears to have been diffracted by the edge of the disk 'C'. With the unobstructed part of the shock wave now all but past the disk (figure 10b), the second shockfront observed in the earlier images 'D' has also become visible. Once again we are able to confirm its position within the shock wave, as being the same as that described earlier for figure 3a,b. The reflected part of the shock wave has also moved on, but has now been all but obscured by the cavitation activity originating on the surface of the disk. However, the reflected part of the shockfront can still be seen leading this part of the wavefront.

5.42 Interaction with a simulation stone

Figures 11a-d show a sequence of individual schlieren images of the incident, transmitted and reflected components of a shock wave, produced during its interaction with a simulation kidney stone. The HMT stones are a mixture of plaster and microballoons, that have compressive and tensile strengths very similar at 11 MPa and 4 MPa respectively, to that reported for kidney stones\(^{15}\). Seen in the images left of centre, with three corners oriented towards the viewer is the stone measuring 30x30x15 mm. Far right tangential to the outer circular border of the schlieren observation area is the partial image of the lens surface. Shock wave propagation in these images is from right to left. In the first two images (figures 11a,b) we see the shock wave prior to and at an intermediate stage, of the interaction in which both the main shockfront and rarefaction component have entered the stone. The reflected
component of the shockfront has emerged, but is obscured by the intensity of the second compression phase just about to impact the surface. Even so, additional detail present in the centre of this component indicates its position and also that at this point in time there is little or no evidence of cavitation activity on the surface. Approximately 4 µs later the region associated with the collapse of the bubbles and the multitude of cavitation transients hits the surface, producing intense cavitation activity (figure 11c). Measurements taken from the images\textsuperscript{16} show this intense region is approximately 11 mm in diameter; however, cavitation activity is not restricted to this small area and can be seen at greater distances from the centre of activity. This intense activity appears to coincide with the emission of a plume of debris. While this link is not altogether confirmed in this image, the next in the sequence (figure 11d) does show debris to the right of the stone. The minute pieces of material are further enhanced since they are apparently providing nucleation sites for 'secondary' cavitation activity. Propagating back towards the lens is the reflected shock wave, which is now seen to be immersed in a sea of cavitation transients emanating from the surface activity. The transmitted part of the shock wave is also seen; however, this has been attenuated during its passage through the stone, resulting in only the main shock and rarefaction components being detected. It is clear from this particular image that the transmitted component of the shock wave has travelled further in the field than the reflected component. By comparing the relative positions of the two components we can determine the sound speed in the simulation material. This also makes it possible to obtain an estimate of the simulation material's acoustic impedance. Analysis of the image (assuming \( c_{\text{water}} = 1480 \text{ ms}^{-1} \)) indicates a sound speed in the material of approximately 2432 \text{ ms}^{-1}, given that the calculated density (wet) was 1097 kgm\(^{-3}\), we find that the material impedance is around \( 2.66 \times 10^6 \text{ kgm}^{-3} \text{s}^{-1} \). Reference to table 5 chapter 2.54, indicates that as far as these parameters are concerned, the material bares more resemblance to a gallstone than to a kidney stone. Pressure measurements taken in the field using the hydrophone have shown that the transmitted component of the shock wave has a peak pressure of \( \approx 1 \text{ MPa} \).

Figure 12a,b shows the front and back surfaces of a simulated kidney stone after the application of 200 shock waves. The shock interaction and cavitation activity at the front face has produced a pit which is essentially conical in shape, 6.5 mm deep, with an approximate base diameter of 11 mm. Additional erosion outside this area is also visible, and is commensurate with the cavitation activity observed in the schlieren images away from the central region. The damage on the rear face is due to spalling, which occurs because of the phase change associated with the reflection from the back face (stone/water interface) of the partially transmitted shock wave (section 2.54). The
Figure 11 (a) The schlieren image of the shock wave before impact. The outline of the model stone is visible 'A' and the acoustic lens on the right 'B'. The region of the acoustic pulse containing cavitation transients is visible in the tail 'C'. (b) The image of the shock wave during the early stages of its interaction with the surface. The first two components of the shock wave have entered the stone. (c) The part of the wave composed of cavitation transients hits the surface, producing additional cavitation events 'd'. While the reflected component of the shock wave 'E' is seen clear of the surface propagating back towards the lens. (d) The surface activity has now subsided as the reflected component of the shock wave and the cavitation transients 'F' propagate back towards the lens. Small pieces of debris can be seen immediately below the reflected shock wave 'G'. The part of the shock wave transmitted through the stone is now visible 'H'. Width of scale bar = 20 mm.

First evidence of spalling was the fracturing of the surface material very early in the shock treatment. Large pieces of material then fell away during the remaining course of the treatment, to reveal further fractures on the new interface. The damage occurs over approximately the same area as that seen on the front surface and is of a similar
area to the transmitted shock component visible in figure 11d. The depth of the damage on the rear surface is not uniform, but is generally around 2.5 mm deep. Assuming that this is also the thickness of the spall or chip removed from the surface, we can determine the duration of the compression phase of the shock wave $\Delta t$ (Equ. 2.02) producing the damage. This was calculated to be approximately 1.68 $\mu$s, which compares favourably with that obtained from hydrophone measurements of the initial compression phase at half width full maximum. All the samples tested showed similar damage after 200 shock waves. While the same damage mechanisms can be expected to be present during the in vitro shock treatment of a real kidney or gallstone, their very different structures result in damage that is different to that described above. The uniform homogenous structure of the simulation material is in contrast to the laminated structure of most real stones. A shock treatment study on a bladder stone has indicated, in this particular case, a preference to fracture at the lamination boundaries. This resulted in the loss of relatively large pieces of material from both the front and back surfaces of the stone. At no time during the study was the characteristic pit, observed with the simulation material, formed. The damage sustained by the simulation stones may therefore not be truly representative of what actually occurs with real stones, however, their regular structure does provide a suitable base for quantitative erosion studies.

A quantitative study of stone erosion undertaken by Granz and Kohler\textsuperscript{17} on the simulation material has indicated that there is a strong linear dependence between the volume eroded and the effective energy of the shock wave at the focus. Their results have been obtained using shock waves generated by an EMAT source, tuned to provide a range of shock wave foci and effective energies. Since the results of their experiments were based on the application of only 40 shock waves, a few of our
samples were also subjected to this amount for direct comparison. As expected, the
damage to the material was less with the reduced number of shock treatments. The
shock treatment produced a pit 2 mm deep with a base diameter of approximately 6
mm, there were also signs of fracturing on the rear surface. Comparison between the
our results and that obtained by Granz and Kohler for the dependence of the eroded
volume on the peak energy density, compared well. Extrapolation of their results for a
volume of 18.84 mm$^3$ indicated an energy density of 0.14 mJ mm$^{-2}$ compared to our
estimated value of 0.106 mJ mm$^{-2}$ (section 3.33). Similarly, the dependence of the
maximum crater depth on the peak energy density yielded a value of 2.08 mm, which
compared well with our result of 2 mm. Extrapolation of the result for the dependence
of the eroded volume on effective energy (where Granz and Kohler define the
effective energy as the energy density multiplied by the base of the crater) yielded a
value of 9.46 mJ for an eroded volume of 18.84 mm$^3$ (section 2.55). In our case this
was calculated to be 3 mJ, a factor of 3 below theirs.

5.43 Discussion

The use of schlieren techniques for the visualisation of shock phenomena provide
distinct advantages over conventional high speed photography. Their use for the study
of the interaction between a shock wave and a simulation stone allows for a clearer
interpretation of events. This is because it is not only possible to visualise semi-
transparent objects such as cavitation bubbles and debris, as with conventional
methods, but also the shock waves which are the key to the whole interaction. The
images have shown that as the lithotripter shock wave passes through the water, the
strong rarefaction component of the wave ruptures the liquid producing cavitation
bubbles. It is not until the tail end of the shock wave that the bubbles collapse to form
spherical cavitation transients, which then form an additional 'shock front'. Only when
this region of the wave is incident upon the stone is any noticeable cavitation activity
seen on the surface.

Sass et al$^1$ have suggested that cavitation occurring within liquid filled flaws of the
stones plays a major role in the disintegration process. Certainly the strength of the
cavitation activity seen on the surface makes this likely. The overall level of activity
may be increased due to the fact that stone debris in flaws and damage sites, will
provide nucleation sites for further cavitation activity. The effectiveness of this
mechanism to cause damage, may be further enhanced due to the concerted collapse of
the clusters of cavities near the surface$^{18}$. Normally, damage is associated with the
collapse of single cavities, which emit shock waves when they collapse symmetrically, or liquid jets when they collapse asymmetrically due to their proximity to a surface. The concept of concerted collapse which has been discussed by Hansson and Mørch\(^7\), predicts the damage potential of the individual cavities increase towards the centre of the cluster. The main reason for this is that the energy radiated inwards from the collapse of cavities at the boundary of the cluster is transferred into collapse energy of the other cavities. Jet impacts and shock waves produced due to the collapse of central cavities, would therefore be more erosive than the other cavities in the cluster. This would certainly go a long way to explaining the shape of the pit eroded on the front face of the simulation stone. The damage associated with the asymmetrical collapse of bubbles and the formation of liquid jets is difficult to assess because the mass of cavitation activity near the surface obscures the fine detail. Nevertheless, it is fair to assume that this occurs and that both mechanisms damage the stone and contribute to the jetlike displacement of material from the surface.

The assessment of the sound speed in the simulation material from the images and the calculation of the material impedance, allowed us to determine the transmission coefficient of the stone. In theory, neglecting attenuation, we should only expect a 15% reduction in the pressure of the transmitted component. However, hydrophone measurements have indicated a peak pressure of only 1 MPa, a reduction of 95%. The absence of cavitation in this part of the field reinforces this finding. This would therefore suggest that significant attenuation has taken place within the stone. The comparisons between the results obtained by Granz and Kohler and the damage associated with the relatively low shock energy densities used for our study, appear to correlate well. The relationship between maximum crater depth and peak energy density and to a certain extent eroded volume against the peak energy density match our findings. However, in the case of the dependence of the eroded volume on the peak energy density, Granz indicates that there is a poor correlation between the two factors. The similarity between our results and those obtained by Granz for these particular characteristics, are unfortunately not repeated in the results for volume erosion against effective energy, where Granz obtained a high data correlation (section 2.55). The reason for this is possibly due to the high accuracy with which Granz has been able to determine the eroded volume from the digitised plasticine form of the crater. Our calculations have relied solely on using the conventional formula and estimated values for the diameter and depth of the pit. The other possible explanation for this wayward result could be due to our effective energy being near the suggested lower limit for stone erosion.
5.5 Cavitation distortion phenomena

During the course of work associated with the imaging of shock waves produced by an EMAT source, many features of the shock wave have been investigated. The schlieren and shadow images of figure 13a,b show the now recognisable form of a focused shock wave in the focal region of the system. The most interesting feature of both these images and many of the others presented earlier, is the apparent distortion to cavitation bubbles and cavitation shock transients within the shock wave itself\textsuperscript{19}. The first indication of this unusual phenomena can be seen in the bubbles present in the vicinity of the second compression (region A). When compared to other bubbles in the field (regions B & C), bubbles in this part of the field appear aspherical. This is born out by measurements, (table 1) which also show that the bubbles range in size up to a maximum of 0.81 mm in diameter in the region of the acoustic field indicated by 'C'.

<table>
<thead>
<tr>
<th>Region A 2nd compression (mm)</th>
<th>Region B 3rd compression (mm)</th>
<th>Region C (mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major axis 0.54±0.09</td>
<td>0.54±0.04</td>
<td>0.71±0.05</td>
</tr>
<tr>
<td>Minor axis 0.35±0.05</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 The average cavitation bubble diameter measured in specific parts of the field.

A high degree of asymmetry is also apparent in the 'primary' cavitation transients i.e. those generated initially within the shock wave. In some cases they appear flattened on their leading edge and pinched in top and bottom (figure 13a) and others, egg shaped (figure 13b). Those transients observed away from the shock wave, predominantly 'secondary' transients, are symmetrical as expected. An acoustic answer to the distortion exhibited by the 'primary' cavitation transients would require changes in density, temperature, etc. very much larger than those expected to be produced by the shock wave. In some cases it was calculated that the speed of selected parts of the wavefront appeared to differ by more than 300 ms\textsuperscript{-1}. Reference to figure 2 section 3.22, indicates that the speed of sound in the water, even at the focal point of the EMAT where local pressures can exceed 20 MPa, increases by no more than 30-40 ms\textsuperscript{-1} at a particular temperature. Up to a certain point a similar increase in velocity can be expected for an increase in temperature. However, the local rise in temperature associated with the passage of the main shock wave is insignificant, and is therefore even less likely to be a cause of the distortion. One very plausible explanation is that the cavitation bubbles cause the distortion. This is based on the fact that an air water mixture of only 1 part in 10,000 parts water lowers the velocity of sound in the water
Figure 13a The schlieren image of a focused EMAT shock wave and its cavitation field. Width of the scale bar is 10 mm.

Figure 13b The shadow image of a similar shock wave and the surrounding field.
by about 40%. However, this too can be rejected on the basis that the cavitation transients outside the shock wave show no signs of distortion.

It is suggested that the apparent lack of symmetry exhibited by the cavitation bubbles and primary cavitation shock transients is due to a complex optical lensing effect\(^\text{20}\). This effect is brought about by the changes in refractive index associated with the compression and rarefaction cycles of the shock wave. Regions of positive pressure act in the same way as a conventional positive lens, except here the extremes of pressure coupled with the fast rise times produce very strong non symmetrical lenses indeed. Conversely the strong rarefaction component of the shock wave produces what is essentially a negative lensing effect. The temporal profile of the shock wave can therefore be modelled by a sequence of both positive and negative lenses. To investigate this phenomena further, the field was seeded with micro-bubbles by circulating the water in the tank through an impeller type water pump (section 5.1). Ordinarily, cavitation transients were only produced in the tail of the shock wave and in the field immediately behind. By seeding the field, it was hoped that a few of the cavitation bubbles produced by the pump, would collapse and produce transients in the part of the field dominated by the shockfront and the main rarefaction component of the shock wave. Figure 14 shows two such transients, the greater pressures and refractive index gradients that exist in this part of the shock wave have resulted in an increased level of asymmetry in the transients. Generally, the transients appeared elongated in the region dominated by the shock front and flattened in the part of the wave associated with the peak negative pressure. The difference in shape between the transients of figure 13 and those of figure 14 is just associated with the position of the cavitation event within the EMAT shock wave.

To model the cavitation transients, a system of concentric circles of radii 3-14 mm in 1 mm intervals was marked on a perspex sheet and positioned in the object plane of the schlieren system. Figure 15 shows a shadow image of the shock wave and the apparent distortion to the rings as the wave passes in front of them. The parts of the rings not obscured by the shock wave remain undistorted, whilst within the shock wave both the actual and apparent positions of the rings can be seen. The visibility of the undistorted rings is due to scattered light from the edges of the lines that has avoided the inhomogeneity but remained in the optical system. These faint images are particularly helpful since they allow a direct comparison to be made between the true and apparent positions of the rings. A comparison between the distortion seen in parts of the cavitation transients of figure 14 and that of the rings, in figure 15, clearly shows the similarity. This is especially so in the cases where the cavitation transients and the
Chapter 5: Schlieren studies

Figure 14 The schlieren image of a focused shock wave passing through a field seeded with cavitation bubbles. Scale bar = 10 mm.

rings are seen through the compression region of the shockfront and appear elongated by the strong positive lensing effects.

The rather complicated displacement exhibited by parts of the cavitation transients and the system of concentric circles in the field of the shock wave of figures 13, 14 and 15, can be understood more easily using a fine wire grid. Figure 16 shows the displacement of the individual wires on a 250 µm wire grid as the focused shock wave passes in front of the grid. Along with the effects of defocusing, seen in the image as a broadening of individual wires, is the relative displacement of each of the grid wires due to a particular part of the shock wave. The vertical components of the grid have been displaced to the right and left, by the negative and positive gradients (rising and falling edges) of the shock wave. The horizontal components have been displaced either up or down by the radial change in the refractive index from the central axis of the EMAT shock wave. Grid images, such as that shown in figure 16 do not lend themselves easily to simple analysis. However, by visually superimposing the shock waves and the distorted cavitation of figures 13 and 14 onto the grid image of figure 16, a clearer understanding of why the cavitation appears distorted can be obtained.
Figure 15 Modelling the cavitation distortion with a series of concentric rings.

Figure 16 The schlieren image of a focused shock wave displacing a 250 μm grid.
5.51 Discussion

Schlieren and shadow images of EMAT shock waves in water have shown cavitation transients which are highly distorted, perhaps indicating propagation velocity differences up to 20% of the sound speed. Schlieren methods used in conjunction with systems of concentric circles, lines or grids have proved to be very useful for understanding the cavitation distortion and in addition the shock wave structure itself. Modelling of the transients with a system of concentric circles has shown qualitatively that the apparent velocity differences seen in parts of the cavitation transients is an optical effect. Comparison of the grid line displacement for a focused shock wave, (figure 16) with the images of normal and seeded cavitation (figures 13 & 14 respectively) shows extremely clearly why the cavitation transients appear distorted in the way they do. A transient straddling a region of compression will appear elongated, whilst the same transient straddling a region of rarefaction, will become compressed and appear flattened. The level of distortion is simply dependent on the strength of the refractive index gradient in that region. Generally, this means that cavitation in the region of the main shock front will appear highly distorted, whilst, the relatively low pressure in the tail of the shock wave results in only a small distortion. This is only generally true, because the strength of the field in the region of the focused shock front fall off rapidly with radial distance away from the centre. Therefore some transients that appear to be near the shock front, such as the larger of the two cavitation transients in figure 16, are in fact away from the central region. In such instances these may only appear as distorted as a transient in the central region of the shock wave tail as seen in figure 13. A quantitative assessment of the cavitation transients from the grid displacement was not possible, because of the masking effect of the shock wave and other transients and because of the relative uncertainty about the position of the transients within the wavefront.

References


CHAPTER 6

CONCLUSION AND SUGGESTIONS FOR FURTHER WORK

6.1 Conclusion

The work detailed in this thesis has been born from the research group’s general interest in imaging cavitation phenomena, particularly that associated with laser interactions in liquids. The desire to apply this knowledge to a practical application, brought about the interest in lithotripsy. The main thrust of the present research has been in the generation and imaging of the shock waves used in the treatment of stone disease. As a consequence of the research in this area, a fascinating insight into the history of the stone and its treatment throughout the ages has been gained. It is known that the treatment of this disease has been practised since very early times and that it may be associated with the first true operation for the relief of a specific surgical condition. Its pedigree does not stop there, it has been speculated that one of the symptoms of the disease could even be an explanation for the first of the biblical plagues, even if this is not the case, it has certainly had some famous victims. Until the advent of anaesthetic and antiseptics c1870, operations to remove stones were invariably fatal or at the very least left the patient with serious side effects.

Today stone disease is known to be responsible for many complaints and cause considerable suffering. It is estimated that worldwide in 1986, £2 billion was spent on the treatment of the disease. The incidence of kidney stones throughout the world appears to be increasing, whilst the incidence of bladder stones is actually decreasing. Men are particularly at risk especially in certain parts of the world, this latter fact suggests that diet is a contributing factor. One such component of the diet, calcium, has for a long time been suspected of increasing the risk of kidney stones, but recent studies suggest it may now be beneficial. The incidence of gallstones shows a similar trend to that of kidney stones, especially in the industrialised west. Once again diet is strongly suspected, with cholesterol being the main concern, although it has been suggested that a person may be genetically susceptible, particularly women and American Indians. Prior to 1982 the most prevalent form of treatment for both kidney and gallstones was their surgical removal. This meant that the patient often required long periods of convalescence. The treatment itself is not without risk, in the case of
The single most important component of a lithotripter system is the shock wave source, as such it is responsible for the major performance differences between the machines. Recent studies suggest that it is total energy and not pressure as originally thought, that is the dominant shock wave parameter effecting stone fragmentation. Machines with electromagnetic sources have been shown to perform very well in clinical studies and justifiably much interest is now being shown in the technology. One such source, the EMAT has been described in this thesis. Being a finite amplitude emitter, the EMAT has to rely on the non linear characteristics of the medium to generate the shock like structure in the wavefront. The typical electroacoustic efficiency of these devices is only around 0.03%, maximisation of the acoustic output through the correct choice of diaphragm, coil/diaphragm separation, number of turns etc. is therefore essential. Theory shows that this can be achieved by constructing a coil with a large number of turns, a small radius and the smallest possible coil/diaphragm separation. The limiting factor on the radius of the coil, is of course the reduction in lens focusing efficiency for increasing values of \( \lambda/D \). The choice of diaphragm was found to be particularly important, since it effects both the efficiency and reliability of the device. The material must have a low resistivity and be strong.
and flexible enough to withstand the force exerted on it by the coil without being permanently deformed.

Because the electroacoustic efficiency of this type of device is so low, achieving the sort of pressures and energy density required for lithotripsy through an increase in discharge current is not feasible. An alternative method is to use an acoustic lens or similar focusing system, but it is well to be aware when designing such a system, that aberrations and diffraction can be significant. However, provided a number of criterion are met, it is possible to reduce their effects to an acceptable level. It is typically found that the focal regions of these devices are characterised by a wider pressure distribution in the direction of propagation, than in the transverse direction. Even so, peak pressures of anything between 20-100 MPa and energy densities of up to 0.5 mJ/mm\(^2\) can be obtained. The performance of our EMAT was somewhat lower than that obtained with commercial devices, although it must be said the reason for this was more financial and the need for reliability than due to limitations of the design.

Measurement of the acoustic field generated by the EMAT and for that matter any shock wave source, poses obvious problems. The conventional way of obtaining information about the pressure distribution in an acoustic field, is to use a hydrophone. Hydrophones designed for measuring such acoustic fields, require very specific properties. As well as being capable of withstanding the high pressures found in the focal regions of such sources, they should also have a stable and linear voltage response over a large dynamic range of pressures and a broad bandwidth with a smooth frequency response. Using a piezoelectric hydrophone designed for optimal impulse response i.e. a wide frequency and compact time domain response, it has been possible to characterise both the planar and focused acoustic fields generated by the EMAT and compare them with the theoretical predictions. The excellent results obtained with such a simple and compact device are due in part to the exceptional properties of the piezopolymer PVDF. However, the long term stability of the hydrophone, particularly its sensitivity, can be expected to gradually drift downwards. The main cause of this is thought to be due to the delamination of the elastomer electrodes, as a result of the stresses set up in the PVDF by the shock wave and possibly even cavitation activity. This loss of sensitivity in devices using PVDF for similar measurements, is very common and has been reported widely. The calibration data for the hydrophone, considering it theoretical characteristics are somewhat dubious. The indications are that the rapid fall in sensitivity above 500 kHz is mainly due to the hydrophone only measuring the spatially averaged pressure in the test field.
Optical techniques for investigating acoustic disturbances have several advantages over conventional hydrophone measurements. They are non-invasive, have a wide field of view and do not suffer the limited time response of some types of hydrophone. Techniques such as schlieren methods rely on the changes in refractive index of a medium that occur when its density is changed due to say compression. The relation between density and the refractive index can be expressed by the Gladstone-Dale relation, which has been shown to produce acceptable results for the compression of water up to 500 MPa. The schlieren methods used to examine the shock waves generated by the EMAT, have yielded both qualitative and quantitative information about the acoustic field. As a result it has been possible to investigate some of the interesting processes that are suspected of being present during lithotripsy. This is particularly so for the sensitive schlieren images, where cavitation associated with the rupture of the water by the shock wave has been identified. The initial or primary cavitation events are restricted to the central region about the axis of the EMAT and particularly in the focal region, where the pressures are the highest. As the shock transients produced by these primary events expand, they excite other cavities present in the field producing secondary cavitation transients. These are seen to fan out and trail behind the shock wave and fill the field with a mass of cavitation activity. The forward propagating components of the primary transients are seen to bunch up in the region of the shock wave associated with the collapse, to form an additional shock region i.e. one composed entirely of cavitation shock transients. As for the shock wave, a comparison between the planar and focused images has shown that the lens adds additional structure to the already complex images of the shock wave, this makes interpretation more difficult.

The problem of interpreting these very complex images has been partially solved by the use of quantitative schlieren methods. Using these methods it has been possible to relate specific components within the schlieren image to the temporal signature of the shock wave. An assessment of the shock wave pressure has also been made possible, particularly with the scale method, where the shock wave has been partially reconstructed from the line displacement data. However, on a note of caution, these methods are only sensitive to the total ray displacement and not the behaviour of a ray in a particular part of the wavefront. Consequently, with complex wavefronts such as those associated with focused shock waves an assessment of pressure can be quite difficult and often very inaccurate. One of the study areas to show the most promise, looked at the interaction between the shock wave and a simulation kidney stone. The studies revealed the incident, transmitted and reflected components of the shock wave and the cavitation activity on the front surface of the stone. This surface cavitation...
appeared to coincide with the arrival of the shock region associated with the cavitation transients, but as yet its contribution, if any, to the surface activity is not known. In common with many other studies, the simulation stones showed signs of cavitation and spalling damage. The possibility of liquid jet damage, which is known to occur, was also not ruled out, although at present no direct observation of the phenomena or the damage associated with its presence have been carried out.

The most interesting finding of the schlieren studies has been the apparent distortion to the shape of cavitation present in the field of the shock wave. This affected both the bubbles and transients, although the effects were more noticeable in the case of the transients. By modelling the transients using the scale method, it has been found that the apparent lack of symmetry exhibited by the cavitation is due to an optical lensing effect. This effect is brought about by the changes in refractive index associated with the compression and rarefaction cycles of the shock wave. Regions of positive pressure act in the same way as a conventional positive lens, except here the extremes of pressure coupled with the very fast rise times produce very strong nonsymmetrical lenses indeed. Conversely the strong rarefaction component of the shock wave produces what is essentially a negative lensing effect.

6.2 Suggestions for further work

The work presented in this thesis has looked at the generation, focusing and imaging of shock waves produced by an EMAT. With such a broad base to the work, many areas particularly those associated with the schlieren studies of the shock waves still require further study. It is expected that the most significant results will be obtained from the interaction work. At present it has only been possible to study a set of individual and unrelated images of the interaction, making it extremely difficult to assess the strength and contribution of the various mechanisms associated with stone fragmentation. Obtaining a sequence of related images of the same event would help the analysis a great deal, but would necessitate replacing the camera with a multi-framing image capture facility. Other modifications to the system might include up rating the output of the EMAT, to nearer that of a commercial lithotripter. An improved lens design would almost certainly assist in achieving this particular goal. Long term, the integration of a Mach-Zehnder interferometer would provide highly accurate fringe shift information about the acoustic fields of these devices. This will be essential if a clearer understanding of the benefits and possible dangers of treatment using lithotripters is to be achieved.
Appendix I Evaluation of the shock wave intensity

From chapter 3 the intensity of a shock wave of pressure \( P \) is given by,

\[
I(r) = \frac{1}{\rho c} \int_0^t P^2(r,t) dt
\]  
(1)

substituting in the expression for the pressure \( P(t) \) (Equ. 3.36), where

\[
P(t) = 2P_m e^{-\alpha t} \cos(\omega t + \frac{\pi}{3})K(1-e^{-\beta t})
\]  
(2)

therefore

\[
(P(t))^2 = 4P_m^2 e^{-2\alpha t} \cos^2(\omega t + \frac{\pi}{3})K^2(1-e^{-\beta t})^2
\]  
(3)

let the constants \( 4P_m^2 K^2 = A \)

Expanding the expression and multiplying out by \( e^{-2\alpha t} \) gives,

\[
I = \frac{1}{\rho c} \int_0^t \cos^2(\omega t + \frac{\pi}{3}) (e^{-2\alpha t} - 2e^{-(\beta+2\alpha)t} + e^{-2(\beta+\alpha)t}) dt
\]  
(4)

(a) (b) (c) (d)

Using Maple on the above expression each of the components a, b, c, d were evaluated. Evaluation of the integral for components a & b gives,

\[
\frac{-1}{4} \frac{1}{(\alpha^2 + \omega^2)\alpha} e^{-2\alpha t} \left[ -\omega \alpha \cos(2\omega t + \frac{\pi}{6}) - \alpha^2 \sin(2\omega t + \frac{\pi}{6}) \right]
\]  
(5)

similarly the evaluation of the integrals for components a & c gives

\[
e^{-t(\beta+2\alpha)} \left[ \cos \frac{\pi}{1}(-2\omega\beta - 4\omega\alpha) - \sin \frac{\pi}{1}(\beta^2 + 4\beta\alpha + 4\alpha^2) + \beta \frac{3}{6} + 4\beta\alpha + 4\alpha^2 + 4\omega^2 \right]
\]  
(6)

\[
\frac{\beta^3 + 6\beta^2\alpha + 12(\beta\alpha)^2 + 8\alpha^3 + 4\omega^2\beta + 8\alpha\omega^2}{\beta^3 + 6\beta^2\alpha + 12(\beta\alpha)^2 + 8\alpha^3 + 4\omega^2\beta + 8\alpha\omega^2}
\]

\[
\%1 = 2\omega t + \frac{\pi}{6}
\]

and components a & d give
The Intensity \( I \) of the shock wave is therefore given by

\[
I = \frac{1}{\rho c} \mathcal{A}[(5) + (6) + (7)]
\]  

(8)

Appendix II Sign convention for thick lenses

Figure 1 shows the sign convention used in the analysis and design of the bi-concave acoustic lens in chapter 3.

**Figure 1 Sign convention.**

- **D** = Lens diameter
- **\( r_1 \)** = Radius of curvature of 1st surface (positive if center of curvature is to the right)
- **\( r_2 \)** = Radius of curvature of 2nd surface (negative if center of curvature is to the left)
- **\( A_1 H \)** = Positive when the first principle point \( H \) is to the right of vertex \( A_1 \).
- **\( A_2 H'' \)** = Positive when the secondary principle point \( H'' \) is to the right of vertex \( A_2 \).
- **s** = Object distance, positive for object (whether real or virtual) to the left of principle point \( H \).
s" = Image distance (s and s" are collectively called conjugate distances, with object and image in conjugate planes), positive for image (whether real or virtual) to the right of the principle point H".
d = Center thickness.
t_e = Edge thickness.
f = Effective focal length (EFL), may be positive (as shown) or negative. f represents both FH and F"H", assuming lens to be surrounded by medium of index 1.

θ = Slope angles are positive when the axis must be rotated counterclockwise through an angle of less than 90 to bring it into coincidence with the ray.

φ = Angles of incidence and refraction are positive when the radius of the surface must be rotated counterclockwise through an angle of less than 90 to bring it into coincidence with the ray.

Appendix III Ray deflection due to a refractive index gradient

The relationship between the refractive index gradient in an inhomogenous medium and the deflection of a light ray can be found using Huygens principle. Figure 2 shows the coordinate system O in which the x-axis represents the direction of increasing refractive index n and the z-axis, the direction of the light ray in which n is assumed constant (2-Dimentional case). According to Huygens theory of elementary waves, let A and B be two points on the same wave surface. The refractive index at A is n and at B is n + dn. By drawing the secondary waves from A to A₁ and from B to B₁ the wave surface is shown to move from the position AB to A₁B₁. In doing so it can be seen that it also turns an angle Δθ towards the higher refractive index at B. θ is the angle between the surface AB and the x-axis and at the same time the angle of inclination of the light ray towards the z-axis.

![Figure 2](image)

Figure 2 The bending of a light ray by a refractive index gradient.
In order to obtain the characteristic formulae for light curvature, the perpendicular BP is drawn to AA'. The angle ABP equals \( \Delta \theta \). Let the difference of height corresponding to the change \( dn \) in refractive index be \( dx = BR \), and \( \Delta x \) and \( \Delta z \) the differentials for the propagation of light in the x and z directions respectively.

If \( AA' \) and \( BB' \) represent the same optical path \( \Delta L \). Then,

\[
\frac{AA'}{n} = \frac{\Delta L}{n} \quad \text{and} \quad \frac{BB'}{n + dn} = \frac{\Delta L}{n + dn}
\]

(1)

Assuming small angles,

\[
\Delta \theta = \frac{AP}{AB}, \quad \frac{AP}{AB} = \frac{AA' - BB'}{n} = \frac{dx}{\cos \theta}
\]

(2)

then

\[
\Delta \theta = \frac{\Delta L}{n} \frac{1}{n} \frac{dn}{dx} \cos \theta
\]

(3)

Since \( QAA' = \theta + \Delta \theta \), then

\[
\Delta z = AA' \cos(\theta + \Delta \theta) = \frac{\Delta L}{n} \cos \theta
\]

(4)

substituting (4) into (3) therefore gives

\[
\Delta \theta = \frac{1}{n} \frac{dn}{dx} \Delta z
\]

(5)

Thus the total angular deflection \( \varepsilon_x' \) in the x-z plane is given by

\[
\varepsilon'_x = \int_{z_1}^{z_2} \frac{1}{n} \frac{dn}{dx} dz
\]

(6)

where \( z_1, z_2 \) are the entrance and exit points of the ray in the inhomogeneity. On leaving the inhomogeneity and entering the air surrounding the water tank, the light ray will be further deflected, such that

\[
n \sin \varepsilon' = n_o \sin \varepsilon
\]

(7)
where $n_o$ is the refractive index of the air surrounding the tank. The final deflection $\varepsilon$ in the $x$-$z$ plane is therefore given by,

$$\varepsilon_x = \frac{1}{n_o} \int_{z_1}^{z_2} \frac{dn}{dx} \, dz \tag{8}$$

In a similar manner it can be shown that the deflection in the $y$-$z$ plane is given by

$$\varepsilon_y = \frac{1}{n_o} \int_{z_1}^{z_2} \frac{dn}{dy} \, dz \tag{9}$$

If the refractive index gradient $dn/dx$ in the inhomogeneity is constant along a path of length $a$, then (8) can be written thus,

$$\varepsilon_x = \frac{a \, dn}{n_o \, dx} \tag{10}$$

Where $a = z_2 - z_1$. 
Appendix IV Publication list


Poster papers

Oral presentations

(1) The application of optical diagnostics to a high energy electromagnetic acoustic transducer. Physics & Technology of Medical Ultrasound Meeting, University of York. 5-6 April 1993.

(2) Pulsed laser shadow and Schlieren photography of a lithotripter shock device. Association for High Speed Photography, University of Wales Cardiff 19-21 July 1993.

(3) Observation of the interaction between a Lithotripter shock wave and a simulated kidney stone. CLEO, Amsterdam, Netherlands, 28 Aug-2 Sep 1994.